UNIVERSITY OF LIVERPOOL

Retail Research in the Age of Big Data: Guiding the Search for Answers

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Sam Comber

August 2021

Contents

Abstract

Consumption binds the worlds of production and marketing to possession and ownership, satisfying the consumer's functional wants and needs, but also fulfilling hedonic desires related to self-gratification and recreational aspiration. Across the last fifty years, growing consumer individualisation through increased personal incomes, education and political empowerment has transformed the importance of consumption spaces, which have further systematized changes in urban growth dynamics. The preservation of attractive, thriving and cohesive retail systems embedded within the urban fabric are now increasingly coupled to urban sustainability. For this reason, building understandings of retail environments has never been more pertinent, with many amenity-led theories of urban development premised on the distribution of place-specific assets like consumption spaces.

A rich history of theories that, for example, describe how urban spatial structure relates configurations of people, households and firms to clusters of attractive consumption environments have guided our understandings of these systems. Yet, despite this rich history, establishing theories that describe retail environments have often relied on coarse approximations of the phenomena under study. While theories of urban consumption behaviour are predicated at a fine spatial granularity (often describing consumer- or store-level activities), their representativeness are often constrained by the misgiving of coarse underlying data. This emanated from the traditional difficulties of acquiring highly-detailed data required to support the design of research hypotheses that describe consumption spaces.

Most recently, the ability to glance into the inner workings of urban systems like consumption spaces has grown exponentially through the increased digitisation of retail transactions, many of which were traditionally fulfilled *offline*. Vast quantities of data that reflect many aspects of human behaviour have recently emerged from three main sources: open datasets exposed by public organisations; data emergent as a by-product of companies moving their business offerings online; and data streams produced by sensor technologies like smartphones and tablets. This transformation of the *data landscape* accessible to urban researchers has increased the volume and diversity of available data sources, enabling the design of research questions at a degree of scale previously unthinkable. Unlike traditional sources of data like censuses or economic surveys, however, these new forms of data were not produced in view of research or policy analysis. Their usefulness emanates from an accidental nature, but a major flaw of this new data pertains to the quality, degree of completeness and unsuitability to traditional statistical techniques. Emphasising the latter point, new forms of semi-structured and unstructured data (such as street-level imagery

and digitised text documents) require techniques borrowed from outside the field of retail geography. Methods that have emerged from significant innovation within the fields of computer science and data mining have been identified as ripe for potential cross-pollination to research problems in retail geography, but require non-trivial programming skills to access. Therefore, researchers equipped with coding ability to analyse non-traditional datasets with untraditional methods are uniquely placed to explore urban phenomena through a much more granular lens than what has been used previously to drive the construction of theoretical premises.

In this thesis, we embrace a data sharing partnership with the Local Data Company that provides unprecedented access to characteristics describing consumption spaces. Crucially, the granularity of this data allows us to explore novel means of empirically testing theories and hypotheses that sit between the intersection of retail geography and urban economics. With this in mind, the central aim of this thesis pairs data access with modern analytical tools to validate, reinterpret and, where appropriate, provide new answers to long-standing theories that explore consumer perceptions of retail environments.

To provide a comprehensive framework for these aims, the thesis initially provides an in-depth review of the theoretical and methodological underpinnings of traditional works exploring consumer spatial behaviour, and how emergent trends in online retail are reshaping it. Alongside this, we critically examine how the evolution of emergent tools and techniques that have permeated areas of urban science have any potential for cross-pollination in retail geographic research. Having reviewed the contextual landscape of consumer spatial behavioural research, we then distil our central aim and problem statement into four research questions, which constitute a series of empirical chapters.

In the first empirical chapter, we assess the extent machine learning methods can be leveraged to enrich data quality, allowing us to integrate different data sources for describing consumption spaces with a rich set of linked attributes. The chapter frames our study as a record linkage exercise, and we borrow innovative natural language processing (NLP) to match the two databases based on attributes encoded in the text representations of commercial addresses. Our proposed method yields successful classification performance, with a precision of 95.5% and recall of 90.2%, indicating the promising potential of machine learning innovations for creating enriched data that can be used in downstream analytical to better understand consumption spaces.

Our second empirical chapter addresses a research question that asks whether urban hierarchies reflect spatial configurations of attractive retail agglomerations. Hierarchies of consumption spaces have long been examined under a series of public and commercial indicators, yet existing works either lack spatial granularity and national-level representation, or exist as commercial products. The contribution estimates retail centre willingness-to-pay (RWTP) scores for retail agglomerations across England and Wales, and their rankings unpack positions which relate to the size, attractiveness and gravity of their composite retailers influence. Using a validation exercise, we show associations between these rankings and characteristics associated with prospering and thriving locations, which further shows a direct validation of Edward Glaeser's theory of consumer cities.

The third empirical chapter evaluates whether visual-only features extracted from images of retail environments reflect different urban consumer experiences. Visual properties such as colour, brightness, shapes and sizes are all features within retail spaces that have been shown to influence consumption behaviour and are important factors when choosing between competing destinations. This contribution trains a deep learning network, a convolutional autoencoder, on a national corpus of storefront images. Visual classifications are then constructed by clustering the representation learnt from the deep learning model. Variables describing the economic health, composition, size and function, and socio-economic characteristics of each individual premise are finally introduced to differentiate between the clusters. Our approach found distinct groupings from the exercise, which implied empirical support for theories that elucidate relationships between visual-only features of retail spaces and different urban consumption uses.

In the fourth empirical chapter, we evaluate which physical characteristics of shopping environments drive the attractiveness of consumption spaces. Finding which physical characteristics represent benefits (or costs) that internalise into location value is of critical importance to retail planners, and a rich set of theories have emerged more generally in urban environments to describe the direction of particular relationships. The contribution links objects from storefront images predicted from a convolutional neural network to subjective preference through an econometric statistical analysis. We use the estimated RWTP scores from the second empirical chapter as a proxy for attractiveness, and find various detected characteristics, such as motor traffic and pedestrian-features, are strong correlates of this.

Through accumulation of these four empirical chapters, this thesis presents a data-driven framework of analysis that pairs unprecedented coverage of consumption spaces with modern analytical tools. In doing so, this work carries practical (and novel) significance in introducing powerful machine learning algorithms from far ranging domains, including NLP and computer vision, to research questions that explore urban consumption behaviour. Of most significance however, the thesis brings new answers and reinterpretations to existing theories that explore consumer perceptions of retail environments using data-driven methodologies. Ultimately, the thesis demonstrates that as the age of big data expands the availability of high quality, granular data describing consumption spaces, retail research that use these sources to explore urban consumption behaviour have a promising future.

Acknowledgements

To begin, I would like to express wholehearted gratitude to my primary academic supervisor and friend, Dr Dani Arribas-Bel, whose unrelenting support and inspiration guided me through the many testing hours of this process. I would also like to extend this thanks to my secondary supervisors, Professor Alex Singleton and Dr Les Dolega. I am indebted to all three for giving up their time and offering their expertise on the various projects we collaborated on across these years.

I would also like to thank existing colleagues and alumni of the Geographic Data Science Lab, whom I shared countless discussions exchanging ideas, advice and support with since first walking into the Roxby building in the Autumn of 2016.

I would like to express my gratitude to the Economic and Social Research Council and Local Data Company who funded my work, without whom this PhD would have remained unwritten. Moreover, I would like to thank the Consumer Data Research Centre for facilitating access to datasets that allowed me to draft the series of stimulating research questions tackled within this thesis.

Lastly, I would like to thank and dedicate this thesis to my family for their resolute support throughout this process.

Declaration of own work

I confirm that this thesis is my own work, is not copied from any other person's thesis (published or unpublished), and has not previously submitted for assessment either at the University of Liverpool or elsewhere. I declare the research articles published in the empirical chapters were directed and led by myself, and the co-authors included in these articles were involved in a supportive manner throughout the research process.

List of Figures

List of Tables

1 — Introduction

1.1 Introduction to research problem

Consumption is defined as spending for the acquisition of goods and utility (Black et al., 2017). As a form of consumption, shopping binds the world of production and marketing to possession and ownership, with this circulation of goods typically conducted within the public realm. When shoppers reveal preferences at the point of sale this represents a public expression of consumption choice behaviour (Lunt and Livingstone, 1992). A consumer undertaking shopping endeavours has acquisitive desires that are tempered by personal control and financial constraint, while the retailer attempts to persuade and seduce a patron into purchasing goods and services. This procurement of products not only satisfies the shopper's functional wants and desires, but also fulfils consumer aspirations such as recreation, self-gratification or socialisation (Sheth et al., 1991). By choosing between these values, a consumer is observed as a 'problem solver' who has various needs that are satisfied by (im)material possessions, with shopping the medium by which these needs are obtained (Lunt and Livingstone, 1992). This problem represents a series of decisions distilled to the following questions – what to purchase, how much to spend, and *where* to buy it.

A principal consideration amongst shoppers is the question of which retail environment to patronise, and is contingent on factors such as product offerings, service provision, weather, or distance. The rational model that describes predictive factors of shopping motives at particular locations can be divided into two types. The first type describes utilitarian shopping value which is derived from efficiencies of the shopping process. This means optimising the ratio between the purchase of the right product at the right price and the shopping effort required to obtain it (Teller et al., 2008). The second type, hedonic value, consists of benefits a consumer experiences from the shopping process. This includes feelings of increased arousal, perceived freedom, fantasy fulfilment and escapism (Jin and Sternquist, 2004). Drivers of patronage intention, therefore, comprise a broad range of influences, and consumers typically adopt a holistic approach to their evaluation of consumption spaces. Aside from direct shopping motives, attributes of retail environments themselves exhibit significant customer drawing power within the catchments they service. From a consumer perspective, sets of retail outlets deliver positive shopping externalities through the bundling of shopping opportunity at a single location (Koster et al., 2019). Such arrangements provide enriched shopping experiences compared to those offered by single stores by enabling 'trip-chaining' behaviour and accessibility to broad varieties of shops and entertainment facilities (Kim, 2002). Thus, beyond utility derived by the act of shopping itself, consumers enjoy additional, indirect benefits related to the scale and location of retail environments (Rosiers et al., 2005; Oner, 2017).

More concisely, desirable attributes of attractive retail environments command demand inflows that create significant multiplier effects within cities and urban areas. Vibrancy provided by the presence of a retail cluster bustling with shopping activity fosters increased social interaction in space, and large concentrations of stores within a market area attract visitors from places outside the agglomeration, generating multiplier effects within the local economy (Oner, 2017). For this reason, the retail sector has long been understood as a pivotal driver underlying the resilience and sustainability of urban environments (Mazza and Rydin, 1997). Under this framework, agglomerations of shopping opportunity are perceived as 'quasi-public goods' like schools or health facilities, with higher volumes of rational individuals drawn to reside closer to shopping areas that offer more attractive goods and services compared to those without. This can be understood more intuitively with a hypothetical example. On the housing market, a residential property up for sale adjacent to a dozen retail units that are homogeneous in composition is likely to carry a muted effect on a prospective home-buyer's willingness to pay. These circumstances are contrasted to the higher willingness to pay for a property with identical structural characteristics, but one that is situated nearby a shopping district with diverse leisure and retail offerings. This location premium stands because the co-location of diverse retail, leisure and service amenities exerts a pull-factor for customers and service providers alike. Consumers will travel to patronize shops and purchase merchandise, but also spend their leisure time enjoying additional services such as bars, restaurants and cafes (Oner, 2017).

Quite literally, consumer's *vote with their feet* when choosing a retail destination deemed appropriate for their leisure time consumption. High volumes of patrons draw sales revenues that boost the local economy, with shopping environments that accommodate diverse tenant mixes a focus of the highest share of consumer purchase intention (Ozuduru and Guldmann, 2014). Alongside this, residential locations that offer access to these most attractive consumption opportunities typically elicit premiums that are reflected by a higher willingness to pay for properties in the local housing market (Teller and Elms, 2012). Since the financial crisis of 2007-08, however, this trend has been muddled by the expansion of online and omni-channel shopping, meaning consumers have become far shrewder in their fulfilment of consumption needs (Grewal et al., 2018b). The eminence of primary and secondary shopping locations has faltered in recent years, becoming de-emphasised through substitution of consumption online. These changes have reshaped the retail hierarchy in a tumultuous fashion, and disrupted the uniformity of *what* a retail hierarchy resembles (Dolega et al., 2019). Despite this, *space still matters*. The positive externalities tied to thriving retail agglomerations carry public and private benefits that underpin aspects of urban growth dynamics (Oner, 2017), which we explore further in Section 1.2. For this reason the availability of attractive shopping areas and consumption spaces remains a critical driver for the growth and development of cities and regions, as several seminal articles proclaimed over a decade ago (Glaeser et al., 2001; Clark et al., 2002).

Yet, despite a host of works that proposition an amenity-led theory of urban development, the theories and hypotheses that make sense of the relationship between attractive consumption spaces and urban development are often premised on coarse approximations of the phenomena under study. In some cases, the generalisability of studies in retail geography are limited by an absence of adequate coverage, which limits the extent of truthfulness upheld by existing explanations. Most recently however, the traceability of human life has grown exponentially through diverse digital footprints which, when aggregated, offer a highly granular lens into the inner workings of urban systems (Batty, 2012). Societal change has driven much of this transformation, with the now ubiquitous presence of sensor technologies embedded in mobile devices, the movement of many transactions historically recorded *offline* to the web, and growing open data releases by public and government organisations at varying geographical scales (Arribas-Bel, 2014). The down-scaling of computational processing power and storage to microchips which can be embedded in everyday objects has also been particularly influential to this change, allowing the seamless integration of micro-technology to everyday life (Kandt and Batty, 2020). A by-product of this has been a redefinition of the data landscape available to retail geographers and, more generally, social scientists. The burgeoning field of *urban analytics*, coined as "the core set of tools employed to deal with problems of big data, urban simulation, and geodemographics" has grown alongside this data movement, by exploiting the opportunities it presents therein (Batty, 2019).

A central tenet of urban analytics is the belief that with parallel growth in real-time data and computational power, the possibility to detect patterns and change across everyday life grows simultaneously (Batty et al., 2012; Kandt and Batty, 2020). However, while poised to offer new potential for urban knowledge on systems like consumption spaces, fundamental challenges posed by urban big data remain. Making sense of millions or billions of observations (alongside the high dimensionality of big data) means emphasis has shifted towards the development of machine learning models able to translate derived insight into new urban theory (Kitchin, 2016). Traditional statistical methods were the historic workhorse of retail geography and urban analytics, but were also designed in an era where patterns were typically extracted from small, clean sample sizes with well-behaved statistical properties (Batty et al., 2012). Recent, significant progress made principally within the field of computer science have expanded the methods available to social scientist's for handling and extracting insights from urban big data (Kitchin, 2016). While these changes might arouse temptation to *let the data speak for itself*, theory becomes more critical than ever when interpreting emergent patterns. A shopper choosing to use their retailer loyalty card reflects a conscious negotiation of everyday consumption experience, and this subjective process is identical whether we are seeking patterns in small *or* big data. As a consequence, to transparently identify credible causal domains, a prerequisite of urban big data studies are their rationalisation by underlying theory and hypotheses that underpin urban dynamics (Kandt and Batty, 2020). Inspired by these debates, this thesis borrows a similar framework from urban analytics and introduces these ideas to the field of retail geography. In doing so, the thesis pairs unprecedented data access to retail environments with new analytical tools to reinterpret, reformulate and validate long-standing theories that explore consumer perceptions of retail environments. These aims are formalised as follows:

1. To provide new answers and reinterpretations to long-standing research questions concerning consumer perceptions of shopping environments.

2. To introduce powerful machine learning, natural language processing and computer vision algorithms to research problems in retail geography that automate perceptual qualifications of shopping spaces.

With these broad aims defined, the remainder of this chapter is organised as follows. The following section outlines a brief overview of the theoretical framework employed within this thesis, providing a grounding in which we frame our research questions. Next, Section 1.3 formalises the proposed research aims into a problem statement that breaks down to a series of research questions answered across their respective empirical chapters. Finally, Section 1.4 details the explicit contribution of each chapter, before outlining the organisational structure employed throughout the dissertation.

1.2 Theoretical framework

Traditional approaches rooted from the 1960s viewed urban development as a consequence of the spatial distribution of the four factors of production – land, labour, capital and entrepreneurship (Alonso, 1964; Richardson, 1977). Since then retailing and consumption possibilities have been emphasised as pivotal drivers of the growth and development of cities and regions (Florida, 2008; Glaeser et al., 2001; Oner, 2017). In the past, rising consumer individualisation accrued through increased citizen income, education, and political empowerment carried the development of more complex (and niche) markets characterised by a volatility of tastes (Clark et al., 2002). This new direction systematized changes in urban growth dynamics, causing a relative collapse in the explanatory power of variables that influence the economic base such as transport costs, distance and local labour costs. As a result, these changes shifted the mix of inputs that traditionally located clusters of households, individuals and firms, inspiring a host of works that link the city's ability to provide consumption opportunity through amenities such as shopping districts to understandings of urban growth patterns (Lloyd and Clark, 2001; Dawson, 2013).

While these theories continue to ground our understandings of the ties between retail systems and urban spaces, it can be argued current growth dynamics have been disrupted by a "perfect storm" in retailing across the last decade. Growth in internet shopping coupled with falling disposable incomes since the 2007-08 financial crisis *and* increased business rates for businesses have gravely threatened the vitality and viability of consumption spaces such as town centres (Wrigley and Lambiri, 2014). Moreover, the composition of these spaces themselves have evolved recently, and now accommodate an increasing demand for leisure units and hospitality services (Dolega et al., 2019). The altered form and function of traditional retail spaces has, according to Hallsworth and Coca-Stefaniak (2018), torn apart our previous understanding that dictates "goods and services are locally supplied in line with local demand". This has led Dolega et al. (2019) to argue for a non-hierarchical classification of retail systems, challenging the pattern of 'centrality' that dominated the organisational models of central place and retail-focused spatial interaction theory (Mumford et al., 2020).

Under this structural change to the fabric of physical retail systems, a requirement of research that attempts to study links between urban growth and retail systems is an understanding that hierarchies are "now at best blurred" (Jones and Livingstone, 2018). An increasing polarisation between primary and secondary locations propagated by the impact of internet retailing is now being felt, with an emergent "convenience culture" paving way for consumption spaces that are increasingly diverse in composition. For this reason, the preservation of thriving, cohesive retail systems are now contingent on factors such as the acquisition of diverse, attractive shopping composition alongside desirable supply and demand dynamics. In recent times, these factors have been found to be tightly coupled to urban sustainability and city vitality. This has led to a growing focus from urban public authorities and place marketing exercises to build the attractiveness of retail spaces in line with the diverse consumption needs of the modern consumer (Oner, 2017). When this is performed successfully, the positive shopping externalities tied to a thriving retail agglomeration carries significant potential to drive public and private benefit, which is further internalised into the housing market. This is because a depth and breadth of amenities within shopping areas and consumption spaces attract home-buyers to the city or region hosting these features (Glaeser et al., 2001; Oner, 2017). Consequently, it is often argued an amenity-based theory underlies the demand for residential space (Brueckner et al., 1999), and locations that satisfy the diverse needs of the modern consumer result in higher property values for amenity-rich locations.

In the UK, this shift has become emboldened by changing planning regulations that have relaxed previous laws that disallowed the conversion of commercial to residential premises, which has since increased the flexibility of land usage in response to changing market conditions (Department for Communities and Local Government, 2011). The desirability of urban environments are increasingly linked to the provision of diverse consumer amenities, as individuals are willing to pay premiums for locations offering attractive consumption opportunities. Successful shopping spaces now stretch beyond the provision of traditional retail categories, and include leisure and hospitality services such as restaurants, performance venues and shopping malls (Teller and Elms, 2010; Wrigley and Dolega, 2011; Wrigley and Lambiri, 2014). In terms of growth dynamics, these ideas link to a proposition constructed in a seminal article by Glaeser et al. (2001): that high-amenity cities are typically found to grow and develop faster than low-amenity cities. This is because the provision of attractive consumption spaces is found to be crucial in attracting highly skilled modern workers, who balance economic and lifestyle opportunity in selecting places to live and work (Florida, 2008). This transition of urban growth dynamics toward consumer amenity provision is, therefore, reflected in the comparative popularity of retail systems as a key component underpinning theoretical and empirical urban geographical analyses (Dawson et al., 2008). Ultimately, these perspectives that emphasise the role of thriving consumer spaces in the development of cities and regions motivate the theoretical framework employed within the thesis.

1.2.1 Drivers of successful consumer experiences

The economic importance of retailing to urban growth dynamics has risen in parallel to a growing social significance of shopping. In the UK total value of retail sales in 2019 exceeded £394 billion, with around one third of consumer expenditure allocated to retail and leisure services (RetailEconomics, 2020). Moreover, the UK's year-on-year percentage change in retail sales volume has sustained a consistent increase between 2010 and 2019 (Sabanoglu, 2020). Clearly, these trends continue to highlight a sustained growth in the consumption of retail goods and services, but which factors influence positive consumption experiences, attracting a patron's custom?

As a whole, drivers of patronage intention are a function of a broad range of influences, with consumers expected to adopt a holistic approach to evaluating consumption spaces (Finn and Louviere, 1996; Bell, 1999). Trends illustrative of an increasing desire to shop are often linked to the pursuit of enjoyment and experience, which has reshaped traditional retail zones to become increasingly service-orientated, enriched by a multitude of diverse consumer opportunities (Jin and Sternquist, 2004; Glaeser et al., 2001; Oner, 2017). Most fundamentally, store clusters within urban spaces create positive shopping externalities through 'trip-chaining' behaviour (Wrigley et al., 2009; Lambiri et al., 2017; Koster et al., 2019). Consumers conducting multi-purpose trips benefit from the convenience of reductions in transport and product search costs through the bundling of wants and needs at a single location (Reimers and Clulow, 2009).

Alongside retailing, diverse and successful consumption spaces also offer shoppers the possibility of spending their leisure time eating, drinking or using entertainment facilities, socialising with friends and relatives in coffee shops, or even using non-retail services such as health and beauty, banks and council administration (Teller and Elms, 2012). In addition, spaces that embrace "convenience culture" and provide seamless shopping experiences to technology-savvy customers are increasingly valued by modern consumers. Services like click and collect, for example, are a much sought-after innovation, allowing the consumer to conveniently decide when and where their shopping is assembled (Davies et al., 2019). While we leave a detailed exposition of these factors to Chapter 2, generic attributes of attractive retail agglomerations also include: accessibility and evaluations of convenience (Ruiz et al., 2004; Reimers and Clulow, 2004); tenant mixes that encompass a wide range of possibilities offered by retail and non-retail tenants (Teller and Reutterer, 2008); visual and auditory stimuli that contribute to ambience and atmospherics (Baker et al., 2002; Michon et al., 2005); and, most importantly, product ranges offered in terms of the assortment of goods sold by retail stores alongside desirable price-value ratios of merchandise (Baker et al., 2002; Léo and Philippe, 2002).

From a consumption perspective, shopping areas that offer these attractive retail-related characteristics influence utility derived by prospective home-owners (Nase et al., 2015). Home-buyers bid up property values in neighbourhoods located nearby retail environments with preferred place attributes, and this additional utility will be internalised into the housing market. From an urban policymaker standpoint, these characteristics of retail environments present a private benefit (or cost) to be considered alongside the wider spectrum of social and environmental factors (Bitter and Krause, 2016). Cost-benefits related to consumer environments moderate the desirability of nearby residential space to prospective home-buyers, meaning empirical data describing these attributes can be used to unpack which combinations link to urban growth and development. For example, retail environments hosting diverse tenant mixes alongside agglomerations of private and public services will carry substantial positive influence on variation in housing prices (Oner, 2017). Establishing commonalities between successful consumption spaces can, therefore, be used to strengthen the "public purpose" nature of decision-making that guides private development, public improvement and the selection from policy alternatives (Nasar, 1987; Bitter and Krause, 2016).

Alongside policymakers, empirical data describing consumer choice behaviour also provides opportunity for retailers to create a differential advantage by adjusting traditional marketing channels of place, price and promotion (Bell, 1999). Restructuring to the traditional retailer landscape through diverse factors such as growth in electronic retailing, preferences for convenient "local" top-up shopping, and transformation of shopping areas to leisure plazas (Chalmers et al., 2012; Wrigley et al., 2015; Helm et al., 2020) have provided an increased number of consumption alternatives. This means attributes of shopping environments (e.g. physical attractiveness) provide opportunities for retail managers to differentiate their offerings. A concern among retailers is that choice of retail area precedes store patronage decisions (Bell, 1999), and so understanding drivers of attraction to consumption spaces is of key interest to retail managers interested in eliciting a competitive advantage amongst this restructuring.

1.2.2 Traditional approaches for measuring consumer perception

Traditional insights into the components of attractive retail environments have previously been gathered by studying data collected through a wide variety of means. This includes: in-person audits of shopping areas using surveying teams (Dolega et al., 2016; Dolega and Lord, 2020); manually reviewing massive numbers of photographs that visualise the consumption space (Petermans et al., 2014; Quartier and Vanrie, 2019); eliciting preferences of consumers patronising a given location through survey-based responses (Babakus et al., 2004; Rayburn and Voss, 2013; Brito et al., 2019); or revealing willingness-to-pay (WTP) for particular characteristics by econometric modelling (Hui et al., 2007; Nase et al., 2013). These studies typically isolate a representative mix of locations across a small sample of retail environments, before deriving *stated* or *revealed* preference through quantitative techniques that model consumer choice behaviour.

Briefly, *stated* preferences are elicited by collecting responses that describe respondent opinions towards hypothetical scenarios or observed perceptions of consumption spaces. In retail studies, a closed question format is typically used to determine the relative importance of different shopping environment attributes. This begins by proposing a conceptual framework of hypothesised relationships between perceived marketing mix characteristics and latent constructs such as attractiveness, purchase intention, or willingness to pay more (Teller and Reutterer, 2008; Nikhashemi et al., 2019). Construction of k-point Likert measurement scales (1 = strongly disagree, k = strong agree) as questionnaire items are then used to infer latent factors. Constructs are factored as a multi-item set of questions. 'Retail tenant mix', for example, includes items such as "... has an attractive variety of retail scores" and "... has a large variety of retail stores", while 'overall attractiveness' might be composed of questions like "How does ... meet your expectations" or "How satisfied are you with ..." (Teller and Reutterer, 2008). Causal modelling techniques such as structural equation modelling (SEM) are then applied to test hypothesised associations between these latent constructs. Having estimated a SEM model, interpretation can then begin by presenting a graphical model that displays the magnitude and significance of pathways between, for example, retail tenant mix and overall attractiveness of shopping areas. Owing to their interpretability, eliciting *stated* preferences via questionnaire items have been the workhorse of a large literature of retail studies that explore drivers of attractive consumer experiences (Oppewal and Holyoake, 2004; El–Adly, 2007). Yet, a limitation of these approaches are their susceptibility to issues of external validity relating to small sample sizes, in addition to their reliance on respondent choices regarding *hypothetical* scenarios that may not properly account for behavioural constraints.

Revealed preferences, on the other hand, infer decisions created by observable actions, and so reflect a preferred measure of preference, motivation and behaviour. Examples have traditionally employed econometric models based on hedonic regression techniques that identify determinants of property prices; this gives indication of the WTP, or consumer preference, for particular components of retail environments (Nase et al., 2013). Typically this literature has focused on the residential property sector, with hedonic studies on retail property still embryonic due to confidentiality associated with retail premise transactions (Rosiers et al., 2005). In the retail context, the range of "spatial" and "non-spatial" price determinants included in hedonic studies of retail property are also far wider owing to the dependence on "non-spatial" factors such as retail image and tenant mixes that influence consumer preference (Mejia and Benjamin, 2002). Four major themes that constitute the principal determinants of retail rents are divided into variables reflecting customer drawing power, retail centre design, location and market characteristics (Nase et al., 2013), which we explore to great detail in Chapter 2. While the literature describes these constituent attributes in great detail, existing studies typically account for these across a small number of retail units within a single urban context. In hedonic retail studies, for example, Rosiers et al. (2005) explore retail rent determinants across a sample of 939 observations in Quebec City, Canada, while Hui et al. (2007) investigates a set of 151 stores in Hong Kong, and Nase et al. (2013) extracts 301 commercial property (office and retail) transactions in Belfast, Northern Ireland.

As spatial and temporal dimensions of human (and consumer) activities are highly dynamic (Steenbruggen et al., 2013), finding empirical data that measure consumer preference at a fine spatio-temporal resolution is critical to decision-making in retail management. Amongst a growing amenity-led theory of urban development, the accurate measurement and estimation of consumer perceptions towards particular shopping area characteristics is paramount to understanding behaviour within these spaces (Dawson, 2013). As a brief review of this literature suggests, hedonic investigations that *reveal* the utility-bearing characteristics of retail environments are still at a point of inception. Moreover, survey-based approaches that extract *stated* preference to particular retail environment characteristics are constrained by hypothetical bias, with responses potentially misleading or poorly thought out. What further limits these two approaches are their use of coarse approximations of the clientele under study, meaning their findings are typically difficult to generalise elsewhere due to problems of external validity and infrequency of observational units (Teller and Reutterer, 2008).

As a response, in this thesis we develop quantitative methods that circumvent these previous challenges by providing novel ways of testing theories rooted in retail geography. Retail geography has traditionally operated in a data-poor environment, whereby measurements of reality were expensive and cumbersome to extract. A collapse in the cost required to capture, store and manipulate digital data has moved the field more generally toward data-rich circumstances. This traction is predicated on the belief that traditional analytical methods are unsuited to extracting insight from massive, heterogeneous databases (Miller, 2010). Methods grown principally, but not exclusively, within the computer science community have circumvented the traditional requirement of data that exhibit well-behaved statistical properties such as independence, normality and stationarity. These approaches, typically coined as data mining methods, perform a variety of tasks including classification, clustering and regression that share a common characteristic of scalability (Miller and Han, 2009).

While these tasks are not novel themselves, innovations stem from the application of data-driven algorithms (such as artificial neural networks, support vector machines and decision tree ensembles) to perform them. These changes carry strong potential to reshape inferences and logical reasoning of problems in retail geography and, more generally, social science. Traditional statistical models demand the social scientist specify a model premised on theory, followed by hypotheses testing and finally a revision of theory based upon the findings. In contrast, the number of potential hypotheses unlocked from big data is too expansive to test exhaustively, despite the triviality many of these might present (Miller, 2010). While retail geography *might* be fertile ground for significant cross-pollination with algorithmic innovation, never has it been more critical to accompany the discovery process with domain-specific knowledge. The discipline is fortunate in that it can borrow from a rich history of theory, and in applying these exciting new methods, it is theory that should be used as background knowledge to distil new learnings from the exhaustive set of potential research questions. In the following section we formalise this problem statement, before outlining the research questions answered within this thesis.

1.3 Problem statement and research questions

As exploration of our theoretical framework has shown, amenity-led theories of urban development are premised on the distribution of place-specific assets like consumption spaces that are known to contribute towards the attractiveness of cities and regions (Oner, 2017). Moreover, from a retailer standpoint, benefits of agglomeration economies lie in the reduction of consumer search and uncertainty costs, in addition to increased total sales volumes as a result of the clustering of similar stores (Eppli and Shilling, 1996; Rosiers et al., 2005). A long history of academic literature on retail environments has evolved around these theories of urban spatial structure that relate spatial configurations of people, households and firms to clusters of attractive store locations (Dennis et al., 2002). Yet, despite this rich history, the integration of theory and practice in retail geography has often relied on coarse approximations of the phenomena under study. While theories of urban consumption activities are predicated at a fine spatial granularity (usually describing, for example, consumer- or store-level activities within cities and regions), the degree to which measurements are representative of experiences and behaviours at this individual-level is a question of debate. Our knowledge of consumption environments are premised on a rich suite of theories that demarcate understandings of purpose and function, but their extent of truthfulness is contingent on the empirical conditions to which we describe and explain these processes and events.

Often theories account for social processes with “empirical tests of the plausibility of the narrative" (DiMaggio, 1995), with them primarily evaluated by the degree to which it offers a close fit to empirical data, followed by the richness of descriptions from its account (Colquitt and Zapata-Phelan, 2007). In retail geography researchers typically follow an inductive approach that uses observation points as empirical evidence to create theoretical constructs and propositions through inductive reasoning. This grounded theory involves an iterative procedure of collating data that best represents the phenomena under study (consumer, retail store, etc.), before analysing these findings in order to build theories that describe how actors interpret their daily realities (Suddaby, 2006). Studies then conclude with a set of propositions that encapsulate the resulting theory (Colquitt and Zapata-Phelan, 2007). Yet, these propositions and understandings of retail environments are often premised on coarse approximations of how actors behave. This is because collating highly-detailed data to support research hypotheses was traditionally cost-intensive and limited in the throughput required to reconstruct the empirical conditions of these complex systems.

To reintroduce an observation made earlier in this introduction, the ability to glance into the inner workings of urban systems has recently grown exponentially as a product of human life being increasingly traceable through diverse digital footprints (Batty, 2013). A redefinition of the "data landscape" available to retail geographers has been enabled by many economic transactions that were traditionally fulfilled *offline* being moved into the web. This archival has facilitated an "accidental side-effect" of large volumes of data that reflect many aspects of human behaviour becoming accessible to researchers (Arribas-Bel, 2014). These changes have made possible the testing of theories within retail geography at levels of detail and scope unimaginable until only recently. Problems of data scarcity have now been surpassed by an era of new datasets that describe quantifiable aspects of retail environments with unprecedented detail.

One such avenue of data availability has stemmed from academic-commercial partnerships. Companies whose business models rely upon data intensive products and services are increasingly turning to collaborations with academic partners to create research synergies and knowledge exchanges. In this thesis, we embrace such a partnership with the Local Data Company (LDC), a retail location intelligence company who "physically track every retail and leisure business across the entire country", and whose data powers strategy and decision-making for clients working across retail, leisure, out-of-home media, investment, property and financial services (LDC, 2020a). Our partnership provides us with a national dataset describing attributes of 700,078 retail, leisure and service properties located within shopping environments across the United Kingdom. This rich set of attributes includes variables describing property-level characteristics such as: geo-location, full address text, business function, vacancy status, number of car parking spaces, and even individual photographs visualising the premise's storefront.

Yet, in order to unlock a full picture from characteristics of the LDC dataset, we also require the introduction of scalable methods able to analyse and obtain understandings from this highly granular and complex source of data. Traditional inferential methods were designed to extract insight from scarce, static and clean datasets that were created in a context of limited availability, and so were scientifically sampled and adhered to strict statistical assumptions (Kitchin, 2016). Thus, we introduce new analytical tools that rely on powerful computational algorithms to process and analyse massive datasets for testing theories in retail geography and urban economics. These methods borrow from the domains of computer science, pattern recognition and visual analytics, which have become a significant area of modern research investment in the computer vision and deep learning communities, but have yet to take foothold within retail geography.

To reconcile the discipline, we pair unprecedented data access to shopping spaces with modern analytical tools to bring new answers and reinterpretations to long-standing theories that explore *consumer perceptions of retail environments*. This direction can be viewed as a positive force towards greater integration between these disciplines, with this "interdisciplinary glue" facilitating the potential for further cross-pollinization through new interactions and knowledge exchanges (Arribas-Bel, 2014). Thus, our problem statement can be distilled to the following research questions, with each addressing a different dimension from which we can build understandings of how retail environments are perceived.

1. To what extent can machine learning methods enrich data linkage for increasing understandings of retail environments? *Chapter 3*

2. Do urban hierarchies reflect spatial configurations of attractive consumption spaces and retail agglomerations? *Chapter 4*

3. Do visual-only features extracted from images of retail environments reflect different urban consumer experiences? *Chapter 5*

4. Which physical characteristics of shopping environments drive the attractiveness of consumption spaces? *Chapter 6*

Each of these research questions are addressed in the chapter indicated in italics, with an explicit answer to each given by the conclusion in Chapter 7. By addressing these questions, this dissertation supports the thesis that pairing new forms of data and analytical techniques provides novel means for empirically testing amenity-led explanations of urban development.

1.4 Key contributions and thesis structure

The chapters of this dissertation can be grouped into three parts: Chapters 1 and 2 provide introduction and required background knowledge; Chapters 3-6 give a detailed account of the contributions distilled from the research questions; and Chapter 7 concludes with a discussion of main findings, before delivering an explicit answer to each of our proposed research questions. Overall, the key contributions of this thesis can be summarised by the following passages.

A first contribution focuses on leveraging new analytical techniques rooted in machine learning to link highly granular, detailed datasets that describe attributes of retail properties. Enriching data quality through this linkage exercise integrates otherwise disparate sources of data, allowing researchers to access attributes of both datasets for the exploration of desired hypotheses. Thus, this contribution seeks to resolve text-based linkage between pairs of retail addresses into matches and non-matches. While address matching has a long tradition in the literature, advances in machine learning have yet to be integrated into the workflow, allowing for the potential of significant cross-pollination. In practice, we empirically evaluate the performance of two recent developments in text-based machine learning – conditional random fields and word2vec – that have not been applied to address-based datasets. This contribution represents a linkage exercise pairing a governmental and private business register of retail premises, allowing us to unlock a linked dataset providing premise-level attributes such as non-domestic tax rates or the number of parking spaces. Ultimately, Chapter 3 describes a workflow for linking address-based datasets, and facilitates a data enrichment exercise for investigating research hypotheses relating to consumer spatial behaviour.

Our second research contribution introduces a statistical technique, a Bayesian multilevel model, to derive indicators that describe hierarchies of retail environments across England and Wales. While shopping spaces have long been examined under a series of milestone reviews, there exists little quantitative evidence describing the performance of retail economies at a fine spatial granularity. Our hierarchies are formed by estimating the retail centre willingness to pay (RWTP) of shopping agglomerations, which reveals the implicit price that home-owners attribute towards their local provision of retail, service and leisure opportunities. In doing so, we unpack rankings of retail environments across a national network, the positions of which relate to the size, attractiveness, and gravity of their composite retailers influence. Top-ranked retail centres, for example, typically offer multi-purpose comparison shopping experiences that draw the widest geographical reach on consumers, but also elicit the highest RWTP. To validate our retail hierarchies, we verify whether the RWTP estimates correlate to socio-economic characteristics that describe the residential location surrounding the retail environment. Thus, across an unprecedented national scale that is uniquely enabled by access to highly granular consumption space data, Chapter 4 tests whether urban spatial structure reflects an amenity-led theory of urban development.

The third contribution uses an unsupervised deep learning technique called convolutional autoencoders (CAEs) to explore relations between visual features of shopping areas and functional characteristics of the surrounding urban environment. This approach is motivated by the long-standing hypothesis that visual characteristics of consumption spaces provide sensory cues that influence consumer experiences and behaviours within these spaces. Moreover, such visual judgements allow consumers to draw fine distinctions when evaluating between competing destinations. While previous research links factors such as proximity to urban consumer behaviour, visual characteristics describing the environmental context around retail premises are typically neglected. To investigate this hypothesis, we apply CAEs to extract visual features from storefront images of leisure, service and retail amenities, before partitioning this collection of features into several clusters. Measures describing attributes of the retail environment such as the economic health, composition, size and function and socio-economic characteristics are then introduced to differentiate between the clusters, allowing us to assess which variables are distinctive for particular groupings. By leveraging unsupervised deep learning, the core contribution of Chapter 5 demonstrates that visual characteristics of shopping spaces reflect patterns of urban consumer behaviour.

The fourth contribution assesses how physical characteristics observed within shopping spaces influence consumer preferences for particular retail environments. In this work, a state-of-the-art computer vision technology known as object detection is used to detect instances of footfall, motor vehicles, pedestrian- and motor-orientated features from street-level images of shopping, leisure and service premises across the national extent. By automating this collation of data, compared to existing studies, our approach represents an unprecedented audit of shopping locations, beyond which would be feasible to obtain from a manual review of images or dispatch of surveying teams. An econometric approach is then used to link predicted objects to subjective preference, which allows an evaluation of which physical characteristics are positively or negatively related to the willingness to pay for retail environments. Across an unprecedented coverage of shopping locations, Chapter 6 helps to direct retail geography a step closer towards understanding which characteristics of the physical environment drive consumer preference.

Finally, the organisational structure of this dissertation has been designed such that common parlance, terminology and required conceptual knowledge is introduced and formally defined in Chapter 2. However, the contributions of Chapters 3-6 may be read individually by the domain expert or reader equipped with pre-requisite conceptual and empirical knowledge learnt from Chapter 2. Thus, the interested reader may choose to begin with an introduction to previous research in Chapter 2, or skip to a desired contribution chapter, referring back to formal definitions and conceptual knowledge in Chapter 2 only where necessary.

2 — Background

This chapter introduces conceptual and empirical knowledge from theories that explore ties between urban spatial structure and consumption spaces; this forms a basis of the theoretical framework employed throughout the dissertation. First, classical theories of urban consumer behaviour relevant to spatial patterns and processes of retail activity are discussed. This discussion evaluates the range of adopted theories, highlighting interrelations and significant gaps in current knowledge alongside, where appropriate, suggestions for future work deemed to be of practical significance. Next, drivers of consumer preference for particular consumption spaces highlighted in the academic literature are outlined, including discussion of how recent restructuring to shopping experiences have affected tastes and preference. Following this, we conduct a selected review of existing literature that elicit consumer behavioural patterns by *stated* and *revealed* preferences, providing examples of each before debating the advantages and limitations of these approaches. In the penultimate section, we highlight applications of new analytical techniques rooted in machine learning to research problems in urban science, describing existing approaches that have significant potential for cross-pollination with retail geography. Finally, we concentrate on particular applications of machine learning for reinterpreting, reformulating and validating long-standing theories that explore consumer perceptions of retail environments.

2.1 Consumer spatial behaviour

2.1.1 Central place theory

The distribution of retail activity is integral to the national economies of Western countries. In Britain, purchases of goods and services account for around one third of total consumer expenditure, with the retail sector supporting the employment of 2.9 million individuals in 2019 (RetailEconomics, 2020). Despite the economic importance of retail distribution, the study of urban consumer behaviour traditionally received comparative neglect as a field of academic enquiry compared to subjects with more scholarly connotation (Smith, 1937). This predilection was shared amongst social scientists, notably by geographers and sociologists, until significant advances in theory describing the spatial manifestation of retail location emerged from town planners (Scott, 2007). In these years, growing attention from geographers to central-place theory added considerably to the academic literature on consumer spatial behaviour. Walter Christaller's seminal study *Die zentralen Orte in Süddeutschland* first aroused interest by English-speaking scholars through Ullman (1941), which stimulated a considerable literature around central places. The spatial arrangement of retail environments in the majority of British cities, for example, was traditionally understood in terms of central place principles (Brown, 1992). British planning professionals argued this explained the prominence of top stores and fashion houses in London, and reasoning why smaller places like Durham lacked department stores (O'Brien and Harris, 1991).

The principal motivation of Christaller's theory propositioned that an urban environment – whether city, town or village – exists primarily to provide products and services for the surrounding area (Scott, 2007). Thus, the implications of central-place theory explain regularities in the spatial system of shopping spaces, and that, in normal circumstances, consumers tend to patronize the closest retail centre offering their required products and services (Dawson, 2013). This framework offers a theoretical basis for the development of hierarchical systems of retail centres at inter-urban scales. Larger centres typically contain more extensive market areas with greater specialisation in service provision, therefore providing more establishments of diverse business types. A nested hierarchy sorts lower-order centres into the same sphere of influence of a higher-order centre, with the hexagon argued as the most advantageous shape for market areas (Lösch, 1940). Retail centres of a higher-order provide all the composite products and services of the lower-order centres, together with a discrete group of higher-order functions (see Figure 2.1) (Scott, 2007).

Central-place theory is based on the classical assumption of *homo economicus*, the economic man, where both suppliers and consumers of services are portrayed as agents equipped with perfect information (Ullman, 1941). This information gifts the ability to make economically rational decisions that are pursued optimally to their subjectively-defined ends. For example, while a supplier of retail services arrives at the optimal location decision, consumers embark on making economically rational journeys to consume (Pred, 1966). From the retailer perspective, the optimal location decision is assumed to be contingent on the 'threshold concept', which argues a service will only be provided in a market area that is capable of supporting it at a profit (Berry and Garrison, 1958). This threshold is determined by spatial competition of retail centres supplying the same products and services, below which the supply is no longer possible. Meanwhile, consumers are argued to typically travel to the most proximate retail centre that exists within range of the individual's residential location, thereby minimising the time-cost budget of the journey.

In the past, the weight of evidence from academic enquiry asserted the validity of central-place theory for explaining the spatial arrangement of shopping centres (Scott, 2007). Yet, later findings from empirical research were proven to be at variance with Christaller's conclusions, with the functional complexity, increased market area size and

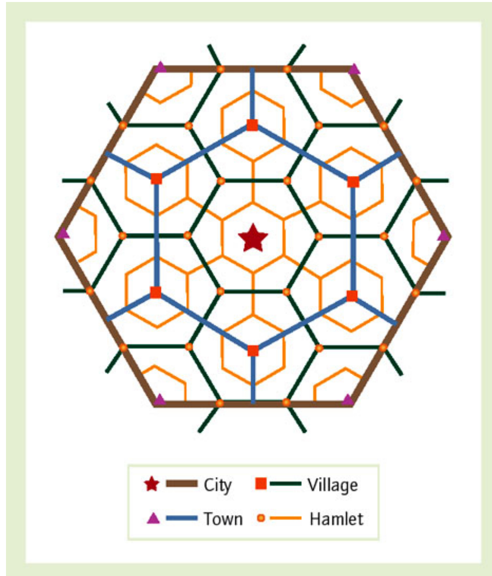


Figure 2.1: Demonstrative example of hierarchical nesting between lower-order and higher-order shopping centres (Warf, 2010).

nested hierarchical pattern found to not always accompany higher-order centres (Kenyon, 1967). Johnston (1966), for example, found that while the metropolitan area of Melbourne, Australia reflected hierarchies of shopping centres, there were significant areal departures in the proportions of each class of centre which deviated from a uniform spatial arrangement. Moreover, consumers, whether resident in metropolitan areas (Clark, 1968) or rural regions (Golledge et al., 1966), have been observed to not necessarily patronize the most proximate centre. A focus on the *nearest centre hypothesis* as the principal behavioural tenet of the theory – whereby a consumer chooses to patronize the nearest centre supplying a product or service – has since been regarded as a significant overstatement of behavioural reality, especially in the Western urban context (Dawson, 2013). For example, a study of the orderliness of consumer spatial behaviour in Christchurch, New Zealand, Clark (1968) found only 50-60% of convenience shopping trips are predicted by the nearest centre hypothesis, with consumers travelling significantly greater distances to the central business district or larger centres.

Various studies highlight deviations from central-place theory where under several circumstances consumer behavioural assumptions diverge from the norm. First, consumers may attempt to minimize transport and search costs by 'trip-chaining' shopping tasks to form a multi-purpose trip, as opposed to minimising travel costs for a single-purpose (Koster et al., 2019). In doing so, shoppers procure both low- and high-order products and services at a higher-order centre located farther away from the nearest lower-order centre (Dawson, 2013). Second, Pred (1966) disputes the classical assumption that consumers are equipped with perfect information. Instead, he argues consumer behaviour is likely to be constrained by incomplete knowledge of retail supply. The economic man assumption requires a consumer possess perfect knowledge of retail facilities, their relative location to competitors, the transport routes to access them, their stock inventories and pricing levels, which insight is self-evidently unrealistic (Downs, 1970). Asch (1952) argues individuals act in terms of what they see, feel and believe, and when mistaken about things, in terms of their erroneous motives, as opposed to things as they are. This will, in addition to social reasons, cause shopping trips to be taken that are not necessarily an optimisation of potential consumption opportunities. Finally, improvements in shopper mobility from smaller towns and cities to larger markets increased visibility to the range of alternative consumption opportunities (Berry, 1963). This increasing accessibility of shopping locations remained a pivotal factor in the emergent overlapping hinterlands of retail centres at all hierarchical levels.

Together, behavioural shopping variation would appear to modify understandings of central-place theory beyond recognition. In Britain, while town residents make *most use* of the nearest retail centre offering a given quality, type and diversity of tenant mix, a multitude of consumer behavioural factors are able to nullify movement minimization (Dawson, 2013). At best, central-place theory offers only a fractional explanation of consumption behaviour in the intra-urban context. While frequently referenced in the literature, the theory concerns only with a particular range of economic activities, precluding goods and services in which demand is dispersed and insensitive to distance (Parr, 2017) – online retailing products that we discuss further in Section 2.2, for example. This is particularly pertinent given dynamic changes underway in retail markets worldwide, guided by growing consumer convenience culture and new technologies carrying transformative action to the physical provision of retail services (Wrigley and Lambiri, 2014). Since the global financial crisis of 2007-08, falling disposable incomes and the expansion of online and omni-channel shopping have meant consumers became shrewder in their fulfilment of consumption needs (Grewal et al., 2018b). These changes have altered the form and function of physical retail spaces, as the role of prime and secondary shopping locations becoming de-emphasised through substitution of consumption activity online. Moreover, Dolega et al. (2019) argues there are no uniformities in approaches for establishing *what* a retail hierarchy resembles. In the UK context, no convincing empirical evidence suggests functional networks of retail centres are hierarchically ordered, with the bulk of this evidence rooted in studies based across the plains of Germany, the US mid-west or centrally managed economies (Parr, 2017).

More recently, the discussion of consumer spatial behaviour has shifted away from central place theory and into the realm of behavioural approaches in the study of shopping behaviour. Thus, while a potentially useful pedagogic tool for understanding fundamental arguments concerning the spatial arrangement of shopping opportunities, the limitations and constraints of central-place theory from its behavioural assumptions suggest alternative research frameworks were required to understand consumer spatial behaviour.

2.1.2 Retail gravitation

Spatial interaction theory proposed an alternative model that incorporated the behavioural constraints that limited the basis in which central-place theory could account for retail hierarchies. The nearest-centre hypothesis was supplemented by the fundamental insight that consumers patronize locations proportional to the attractiveness of retail centres (Joseph and Kuby, 2011). Consumer patronage behaviour was, therefore, assumed to reflect a complex trade-off between the size (or attraction) of retail centres against the friction of distance, or distance decay (Huff and Jenks, 1968). These *gravity models* of consumer attraction were derived from 'the law of retail gravitation', and motivated by analogy to Newtonian physics (Reilly, 1931). This law emerged from William Reilly's ambition to identify break points in the demand surface between various sized cities at different distances (Joseph and Kuby, 2011). Reilly formulated that two cities attract retail trade from an intermediate residential location in proportion to the population sizes of both cities, and in inverse proportion to squared distances between the intermediate town and two cities (Scott, 2007). An early application in Denmark, for example, Illeris (1967) used a gravity model formulation to delimit the hinterlands of central places across various hierarchical levels, validating them alongside hinterlands produced through empirical data of interview surveys from retailers and home-owners.

Yet Reilly's formulation was deemed inappropriately deterministic. This is because assigning the potential sales of trading areas to particular stores or cities and disallowing sales outside that trading area seemed somewhat unrealistic (Joseph and Kuby, 2011). Huff (1964) argued this *all-or-nothing* treatment of trade area delineation reflected an unrealistic representation of consumer behaviour, believing the sales potential of retail centres reflected a gradual decline that could be explained by varying degrees of probability. Thus, Huff (1964) sought a model to account for potential overlapping catchments within a given geographical area. The Huff model addressed these limitations by producing probability isopleths that considered consumer patronage to vary in proportion to the size (or attraction) of the retail centre and inversely proportional to distance from residential areas, in addition to competing opportunities offered by all other centres in the system (Dawson, 2013). More concretely, this simultaneous probabilistic reformulation of Reilly's gravity model was able to estimate the probability a consumer at origin i will shop at retail centre

j.

$$E_{ij} = C_i \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}} \tag{2.1}$$

where,

E_{ij} is the expected number of consumers at point i that are likely to patronize retail centre j;

S_j denotes the size (or proxy for shopping attraction) of centre j. Example measures include square feet of market area dedicated to the sale of products and services;

T_{ij} reflects the distance or travel-time from area i to centre j;

λ is an exponent that is empirically estimated using known origin-destination data to reflect the effect of distance disincentives on various kinds of shopping trips;

C_i denotes the total number of consumption expenditure of residents in area i.

Early applications of Huff models emerged in the 1960s, with Lakshmanan and Hansen (1965) formulating a market-potential model for Baltimore, U.S, finding significant overlap between retail sales generated by the model with actual sales of six large retail centres. More recently, advancement and wider commercial application has been observed for individual grocery stores (Beule et al., 2014) and entire retail agglomerations (Dolega et al., 2016; Sevtsuk and Kalvo, 2018). Methods inspired from Huff have also been used for delineating catchments that model spatial interaction within market areas of retail outlets, demarcating the areal extent from which the main patrons of a store will be found (Davies et al., 2019). Other prominent extensions to the Huff model include the competing destinations model (Fotheringham, 1983) which added an accessibility variable to Equation 2.1, as it assumed the spatial arrangement of destinations would influence trip distribution and, mechanically, patronage of certain locations.

While spatial interaction theory introduced the fundamental insight that consumer patronage is contingent on attractiveness and not simply closest distance (Joseph and Kuby, 2011), determinants of behaviour were premised on aggregate consumer theory (Jensen-Butler, 1972). Recognising these limitations invoked an increased emphasis on individual-level scales of analysis, with even Huff (1960) conceptualising the consumer decision-making process as including a richer set of factors like the breadth of merchandise, number of personal amenities and product values that are difficult to incorporate into retail-based gravity models. Consequently, the absence of these finer drivers of attractive shopping experiences generated considerable problems with the gravity model's ability to build comprehensive understandings of consumer spatial behaviour.

2.1.3 Consumer behavioural approaches

The constraints imposed by the aggregate behavioural assumptions of spatial interaction theory brought an increased focus towards unpacking the nature of, and motivations behind, consumer spatial behaviour (Dawson, 2013). In contrast to deterministic location theory, behavioural approaches heeded specific focus to individual actions, which are a function of environmental context and decision-making processes relative to it (Downs, 1970). Hence, growth in behavioural research of customer motives emerged in two main directions: empirical behavioural approaches that focused explicitly upon classifying aspects such as social stratification and residential location; and cognitive approaches that concentrated on perceptual dimensions of consumer patronage intention. Although both deserve attention, given this dissertation's research focus toward consumer perceptions of shopping environments, in the remainder of this section we critically discuss cognitive variants of traditional consumer behavioural research methods.

The underlying driver of perceptual factors influencing consumer decision-making was argued to be derived from Isard (1956), who propositioned the concept of *individual space preferences*. This notion is premised on perceptions of available alternatives, which is argued as the principal stimulus of shopping behaviour (Dawson, 2013). Perceptual qualifications of retail agglomeration attributes influence the share of spending, time and choices of consumers relative to competing opportunities elsewhere (Teller and Elms, 2010). One of the first comprehensive investigations into the underlying perceptual determinants of consumer spatial behaviour was conducted by Downs (1970), who used individual decision-makers as the basic unit of analysis. Descriptions for the perceptual range of a shopping centre in Bristol, UK were undertaken using questionnaire analysis that specifically excluded the factor of distance to determine how particular attributes of shopping opportunities were likely to be evaluated. Results were analysed using principal component analysis, which found eight factors that described cognitive categories of retail area image, including: service quality; visual appearance; shopping range; and structure and design. In similar vein, several studies later sought to explore 'trip motivations', where insights into the determinants of shopping behaviour were derived from survey respondents asked to indicate stated preferences towards their most important considered features when selecting a retail centre to patronise (Davies, 1973; Bearden, 1977; Bellenger et al., 1977). Under these approaches, the data are typically used in conjunction with stated preference models to identify the most important determinants of consumer choice behaviour[1].

While conceptually simple, numerous studies have analysed survey responses to consistently identify a number of similar factors that drive consumer spatial behaviour. Accessibility, alongside competitiveness and qualitative service characteristics of retail centres, traditionally emerged as the strongest determinants of consumer choice (Dawson, 2013). Elsewhere, more explicit environment-related drivers have been identified as easing or

[1]Given this section's focus on the development of classical theory regarding consumer spatial behaviour, we leave an explicit discussion of stated preference models to Section 2.3.

enriching the the process of consumption, including the ambience of the agglomeration *vis-a-vis* sensual stimuli and atmosphere (Bloch et al., 1994; Wakefield and Baker, 1998) in addition to opportunities that facilitate the potential for social interaction (Wagner, 1975). These attributes relate to the qualitative image of stores envisioned by consumers, which links to a common research focus in the marketing literature of 'store personality'. This line of study has served as a corrective factor to the tendency of geographers to overstate the spatial dimension of consumer perceptions of retail environments. Here, the marketing literature has built consumer perceptions of shopping spaces around non-geographical attributes, embodied by Martineau (1958), who argues that stores have "a total image of many more meanings in the consumer's mind than that of a place for day-to-day transactions." These arguments have sought to clarify the manner in which distance is perceived by patrons, with the suggestion that perceptual aspects of consumer decision-making were increasingly prominent and integral to the study of patronage intention in later years (Dawson, 2013).

2.1.4 Critical reflections and future

As a whole, geographers have played a pre-eminent role in the development of spatial interaction and stated preference models for building understandings of consumer spatial behaviour (Timmermans, 2004). Nevertheless, critical reflection of the assumptions underlying these approaches suggest our understandings of these theories are not always premised on convincing grounds, an issue we return to in greater detail in Section 2.3. Spatial interaction models make simplistic assumptions regarding aggregate consumer behaviour (Jensen-Butler, 1972), while cognitive-behavioural studies are often constrained by small sample sizes due to costs in acquiring survey respondents. Moreover, the traditional instrument of research used to elicit consumer behavioural tendencies, the questionnaire, has been shown as limited by the respondent's modest power of recall (Chrysochou, 2017).

In the past, various accuracy-validation surveys have shown that reports offered by consumers interviewed immediately after a shopping trip are often faulted through omission or mis-sequencing of events (Vanhuele and Drèze, 2002). Direct observations of recording shopping behaviour in the field are also replete with difficulty, being highly labour- and cost-intensive, but also provoking potential ethical objection relating to the non-consensual observation of individual shopper activity for extended periods of time (Dawson, 2013).

Most poignantly, although the development of these methods has traditionally been touted as the best-case line of enquiry, findings are often generalized from coarse approximations that exhibit limited geographic reach. For this reason, reaffirming understandings through an expansion of methods that describe consumer spatial behaviour at finer spatial resolutions represents the next logical advancement in the field. Despite their drawbacks, however, spatial interaction and consumer behavioural approaches remain the workhorse of existing research concerning consumer spatial behaviour. These approaches have shown huge success in revealing dimensions of attractiveness that drive experiences of consumer preference, of which we evaluate in the following section.

2.2 Drivers of shopping preference

Retail agglomerations are described by their marketing mix components, which reflect outcomes of managerially-determined decision parameters and locational factors (Teller and Reutterer, 2008). The array of retail centre characteristics are perceived by potential consumers, who convert perceptions into evaluations of relative attractiveness, which drives behavioural consequences and preferences (Finn and Louviere, 1996). The difficulty of understanding which characteristics drive positive experiences lies in distinguishing whether it is perceptual drivers that cause consumers to choose between alternative shopping opportunity, or whether this spatial behaviour is determined partly by intervening stimuli

(Downs, 1970). In this section, we comprehensively review the eclectic range of factors that drive consumer perceptions of retail environments, which motivates the underlying theory of research outcomes described among the later chapters.

2.2.1 Retail-related attractiveness of urban places

Positive shopping externalities

Empirical findings have shown that a multitude of factors influence the process that drives retail attraction. The consumer's system of perceptions and evaluations towards constructs associated with retail attractiveness are highly interrelated, but most fundamentally relate to the utility derived from the act of shopping (Oner, 2017). By definition, retail agglomerations are sets of stores that provide products and services to consumers that operate within a close proximity retail market. Retail agglomerations represent an important civic function for urban environments, and across European retail real estate have a gross leasable area constituting over 200 million square metres (ICSC, 2017). Arguably, the most important driver of retail attraction is the presence of shopping externalities generated by clusters of stores. Synergistic effects between agglomeration tenants allow 'trip-chaining' behaviour, as consumers visiting several shops during a single trip benefit from increasing returns to scale (Koster et al., 2019). Nelson (1958) first described these agglomeration effects by the 'cumulative attraction' of store clusters, highlighting consumer preference towards spatial retail networks over isolated retail locations. From the consumer point of view, visiting several shops (or undertaking multi-purpose trips) represents a reduction in search and transport costs, which are enhanced when multiple retail stores locate in close proximity (Claycombe, 1991). This is because shopping at retail agglomerations increases the likelihood of finding all required items across a single trip – products comprised of several components like a clothing outfit might require the consumer to visit a range of fashion and footwear stores, for example. Several empirical studies examine behavioural tendencies within retail agglomerations. Using a survey questionnaire, Oppewal and Holyoake (2004), found that consumers are more inclined to purchase bundles of multiple non-substitutable products for a single, combined price when more retail competitors are operating nearby. Similarly, de Palma et al. (2010) also finds that retail agglomerations increase in attractiveness when they minimize transport costs required to hop between stores, which is enabled by trip-chaining behaviour within these spaces.

A discussion concerning the attractiveness of a retail location is, therefore, tied to the richness of consumption possibilities offered within the agglomeration. Retail is progressively viewed as a vital amenity underlying the urban fabric, with *some* data suggesting many consumers in Western countries are increasing their proportion of disposable income allocated to the consumption of leisure and retail services (Oner, 2017). Changing consumer needs and growing interest to the shopping experience has transformed expectations of retail landscapes, with consumer preferences shifting towards consumption spaces and shopping areas that are service and leisure orientated (Wrigley and Lambiri, 2014). These changes suggest the activity of shopping is becoming increasingly linked to the concept of enjoyment and experience (Jin and Sternquist, 2004). A consumer's desire to patronise a shopping destination is, therefore, associated with the fulfilment of needs within a physical environment that can satisfy their leisure and retail desires, in addition to how much value can be derived from the shopping experience at a particular retail centre (Hart et al., 2007).

Aside from purchasing habits, consumers are not always required to spend their money at a given shopping destination to enjoy the visual properties and atmosphere of attractive retail environments (Oner, 2017). The vibrancy elicited by patrons moving around retail agglomerations fosters increased interaction in space, as shopping trips extend beyond simply satisfying a bundle of wants and desires. Consumers travelling to retail clusters are often inclined to enjoy other amenities, and will expend their leisure time eating, drinking and using entertainment facilities with friends and family, in addition to using other services such as banks and non-retail businesses (Teller and Elms, 2010). This presence of consumers walking up and down a given consumption space – the footfall of an area – contributes to its vitality and viability (Mumford et al., 2020), and has historically been used as a key indicator of town centre attractiveness (DoE, 1996). This is because footfall counts are typically associated with the level of attractiveness a location is deemed to possess, alongside its ability to satisfy catchment needs and as an indicator of potential spend (Mumford et al., 2020). Therefore, while location-based factors are most often used to gauge the potential attractiveness for retail spaces in modelling exercises, footfall provides a measure of actual day-to-day visitor patterns, and has the advantage in revealing changes in consumer behaviour and attitudes with lower temporal lag (Dolega et al., 2016). For this reason footfall is often understood as a principal driver of properties such as "town centredness" and frequently underpins the qualities of 'live' consumption spaces (Hillier, 1999; Dolega et al., 2016; Mumford et al., 2020). All together, the above discussion implies not only do retail-related characteristics incite patronage behaviour and underpin the attractiveness of retail clusters, but as do the environments consumers partake in themselves too.

Place-based attributes and value perception

In more tangible terms, shopping tasks are understood as an outcome of a holistic evaluation of perceived place and value attributes. Within the consumer's system of perceptual and evaluative qualifications of retail spaces, a number of marketing mix factors and attractiveness dimensions are proposed to affect patronage behaviour. Teller and Reutterer (2008)'s conceptual framework describing the evaluation process of a retail agglomeration's attractiveness is visualised in Figure 2.2. Every agglomeration attribute and every management decision undertaken by retail managers that influence agglomeration effects can be interpreted as a possible effect driver (Teller and Schnedlitz, 2012). Site-related drivers that relate to the location of urban retail agglomerations reflect the spatial and temporal distances consumers have to overcome between their point of origin – residential or employment location, for example – and the retailer's premises (Teller and Reutterer, 2008). The principal factor describing how a location is reached, *accessibility*, is argued to be composed of three dimensions. These comprise the degree of (in-)convenience regarding the shopping endeavour, the speed of access, and the number of obstacles obstructing the way (Wakefield and Baker, 1998). Owing to the importance of cars for the transportation of merchandise bought while shopping, this includes the connectivity, signage and routing of the road network around the location (Guy, 2007), in addition to parking conditions in and nearby the retail agglomeration which reflects an integrative part of perceived accessibility (Teller and Elms, 2010). This latter characteristic is described by factors such as the availability and cost of parking lots, the variety of parking facilities, and the extent to which the retail environment is accessible from the parking lots itself (Teller and Elms, 2012). While previous literature have also included distance as a separate attribute that links to accessibility (Huff, 1964), Teller and Elms (2012) argue perceptual dimensions of distance are difficult to capture. This is because shoppers typically exhibit difficulty in precisely recalling the logistics of their shopping efforts into units such as minutes and metres, meaning distance is often an inadequate criterion of accessibility (Teller et al., 2006). Overall, accessibility-based drivers that steer consumer preference for particular retail locations are linked to a 'rationalisation effect' that motivates the shopping endeavour in terms of logistical possibilities (Teller and Reutterer, 2008).

The extent to which consumers can satisfy their wants and desires through retail (and non-retail) offerings of shopping destinations represents an additional driver of attractiveness. The variety of composite retailers, number and type of retail and non-retail tenants – cafes, gastronomy or entertainment facilities, for example – within retail agglomerations reflect the range of consumption possibilities, with diverse tenant mixes also minimizing the

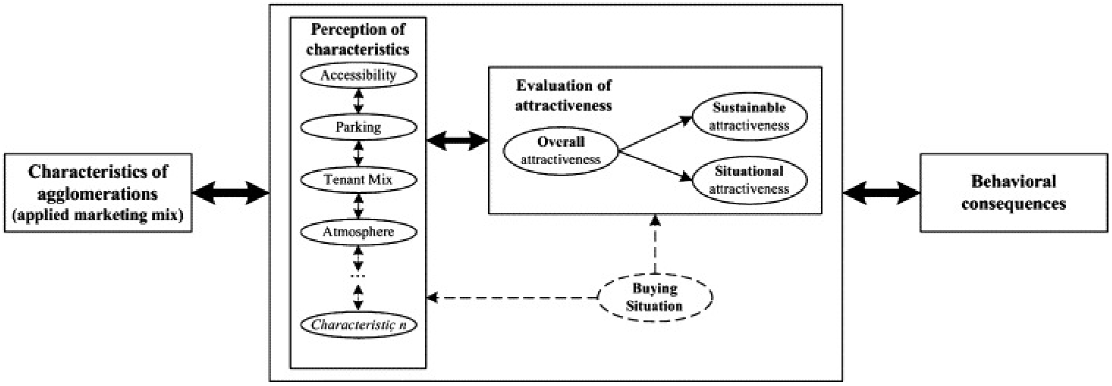


Figure 2.2: Conceptual framework outlining the consumer evaluation process for determining a retail centre's attractiveness (Teller and Reutterer, 2008).

logistics of shopping endeavours (Chebat et al., 2010). According to Teller and Reutterer (2008), a varied mix of tenants has been shown to drive two positive shopping externalities by accumulation and enrichment effects. An 'accumulation effect' reflects the benefits shoppers experience by the enabling of multi-purpose shopping trips which are facilitated by the combined retail and non-retail offerings within the agglomeration. Secondly, an 'enrichment effect' denotes benefits generated through non-retail offerings which supplement the consumer's shopping trip by recreation, entertainment and social interaction. These tenant-related drivers of retail attractiveness also relate to the retail offer promoted within agglomerations. This is based upon value perceptions inferred from *merchandise value*, which is contingent on overall price, quality and price-quality ratio of merchandise, alongside the number of price promotional offers available (Baker et al., 2002). Supplementary to this, the friendliness, competencies and eagerness to assist in the shopping endeavour from sales personnel are also identifiable determinants describing the 'retail face' of shopping destinations, which also potentially attract patrons to consume within these spaces (Teller and Elms, 2012).

Finally, while not as important as those already mentioned, management-related drivers also factor into consumer patronage intentions. This is because they facilitate a comfortable shopping experience that is used to increase the attractiveness for consumers and, by extension, the success of the retail agglomeration and its tenants (Teller and Elms, 2010). This involves the planning, coordination and control of drivers that are highly specific to the retail agglomeration. For example, this might include the provision of public restrooms, ATMs, consumer recreational facilities or even information counters that direct patrons to their desired shopping destination (Baker et al., 2002). Alongside this, the extent of enclosure a shopping environment offers to climatic extremes such as noise, traffic and odours may influence consumer behaviour, with sheltered and pleasant environments encouraging consumers to relax and enjoy the shopping experience (Reimers and Clulow, 2009). Lastly, measures or services controlled by the central agglomeration management such as the employment of security and cleaning services, appropriate store opening hours, consistent branding of the retail agglomeration and a well-designed arrangement and orientation of stores have all been identified as factors that elicit high levels of consumer satisfaction and preference (Severin et al., 2001; Teller and Reutterer, 2008; Teller and Schnedlitz, 2012).

Retail area image

Alongside the aforementioned drivers, consumer preferences have also been argued as relating to the set of visual, auditory and olfactory stimuli confronting consumers that are actively (or passively) used by retail management to ease or enrich the process of purchasing products and services across retail agglomerations (Teller and Elms, 2012). As competitiveness within the retail sector has intensified, managers have sought to transcend a product focus towards enriched customer experiences (Puccinelli et al., 2009). Unsurprisingly, consumers have traditionally been shown to spend more leisure time in those shopping environments they find pleasant (Donovan et al., 1994). Finn and Louviere (1996), for example, argue retail area image perceptions to be a significant determinant of patronage consideration and choice. It is unsurprising an emergence of sensory marketing approaches have began to make sense of the richness and complexity of the customer experience.

To align with consumer preference, products, settings and consumption spaces are increasingly sculpted to appeal to consumers through both rational and emotional levels across multiple senses (Spence et al., 2014). Overall perceptions that define experiences of urban places include feelings of safety, alongside overall perceptions of atmospheric stimuli like temperature, ambience, lightness, air, sound, odour and cleanliness (Baker et al., 2002; Teller and Elms, 2010; Spence et al., 2014). While these stimuli are difficult to control due to the open-air character of retail agglomerations like shopping high streets, the *atmosphere* can be viewed as a unique feature of urban retailing, which might be leveraged to a competitive advantage (Hackett and Foxall, 1994; Teller and Elms, 2012). All elements of retail atmospherics combine to form the consumer experience, which relays the customer's cognitive, affective, emotional, social and physical response to the retailer (Roggeveen et al., 2020). For example, the experience upon entering the premises of UK cosmetics retailer, Lush, overloads the consumer with the pleasant aroma and scent of hand-made creams, soaps, shampoos and other cosmetics for the face, hair and body. This journey is heavily immersive, with sales personnel engaging with customers and encouraging product sampling options. In addition, the store merchandise is beautifully set amongst greenery and rustic store furniture. These sensory marketing approaches generate more touch points with the customer, allowing retailers to differentiate their offering. Spence et al. (2014) formalises this into an organizing framework consisting of several attributes that affect different stages of the customer journey. Crucially, this framework complements rather than replaces the shopping externality and place-based drivers of consumer preference.

The most important domain consists of visual atmospherics which regard items such as colour, brightness, size and shapes of retail spaces that affect encountered levels of stimulation. In the past, Bellizzi and Hite (1992) found consumers preferred blue over red coloured store designs, as blue atmospheres encouraged longer merchandise browsing and greater purchase intention. Similarly, Ward et al. (1992) demonstrate exterior resemblance among fast food restaurants as a significant positive predictor of typicality, attitudes and market share. More recently, Puccinelli et al. (2013) showed male consumers perceived higher savings when prices were presented in red (as opposed to black), while females appeared to show natural inclination towards greater elaboration of the ad and demonstrated greater price recall. Together, studies have shown that visual cues trigger responses in consumers which causes different decision-making behaviour under certain conditions. Changing the visual atmospherics of in-store colour schemes, brightness or hue of lighting conditions are factors that influence purchase intention and sales, and so enter into attitudes of consumer preference for particular consumption spaces (Bell, 1999). Compared to other senses, visual cues have generally been shown to exert a stronger effect on human perception and other sensory stimuli, although other studies have disproved this (Calvert et al., 2004).

While not inherently spatial, additional elements such as auditory atmospherics also contribute significantly toward the consumer liking of shopping spaces. Auditory elements such as volume, pitch, rhythm, tempo and emotional tone of sounds have been shown to cause a positive effect on shopper patronage behaviour (Spence et al., 2014). A meta-analysis undertaken by Garlin and Owen (2006) found the presence of familiar and likeable music carried a positive influence on consumer patronage behaviour as a result of enhancing perceptions of store atmosphere. For example, not long ago London's famous department store, Harrods, installed a reactive and multisensory sound system in its toy department to evoke further positive associations amongst shoppers (Krishna, 2013). These auditory devices are used to convey brand positioning by establishing an appropriate impression on customers browsing the retailer's merchandise. Studies show that constructing *soundscapes* creates immersive experiences that affect the consumer's degree of stimulation (Spence et al., 2014). Musical tempo, for example, has been shown to affect a consumer's perceived passage of time in retail and consumption spaces (Oakes, 2003). For retail managers, this reflects a positive outcome as higher dwell times inside shopping destinations are typically linked to an increased number of items purchased. Thus, by adjusting auditory atmospherics, retailers possess the means to control the flow of customers across the shopper floorspace (Spence et al., 2014).

Lastly, the introduction of appealing fragrances and scents into retail settings may elicit approach behaviour. Biswas and Szocs (2019), for example, show that when consumers smell an indulgent food product for longer than a two minute duration, they purchase healthy products more than unhealthy items, with this effect reversed if the scent is available for less than thirty seconds. Similarly, Madzharov et al. (2015) consider food-related ambient scents that evoke varying perceived temperatures. They show that cinnamon describes a warm scent and peppermint a cool scent, which alter perceptions and choices such that experiencing a warm ambient scent generates increased premium purchases and higher overall spending, in addition to participants perceiving greater social density in the shopping location. Indeed, olfactory cues are often thought to exert the strongest memory recollection of all the senses after first smelling the scent (Goldman and Seamon, 1992), with scented items and shopping environments particularly well recalled (Krishna et al., 2010). Most scent marketing approaches are, therefore, uniquely suited to retailers whose merchandise range includes fragranced products such as soap, chocolate and coffee (Spence et al., 2014). Outside of fragranced products, however, Spangenberg et al. (2006) showed that shoppers are highly likely to exert approach behaviour in the presence of an ambient scent congruent with gender-based items. They found adding a vanilla scent to a women's clothing department and a sweet floral scent to the men's doubled sales in both sections of the store. This demonstrates that retailers can capitalise on scent marketing by identifying an appropriate scent that is pleasing and congruent with the store and brand identity to drive positive evaluations of the store environment, merchandise and sales (Parsons, 2009).

Most research linking store atmospherics to drivers of consumer preference for particular shopping environments focus on single aspects of retail spaces such as sound, lighting or scent. In reality, manipulation of several environmental aspects are required to control the multisensory properties of shopping environments for altering consumer behaviour. Moreover, Roggeveen et al. (2020) argue of the importance for broadening conceptualisations to recognise the customer experience as a longer journey that begins outside physical store environments. Beyond aspects which take place within stores, stages of the customer journey also take place online, where myriad retail touchpoints are established to drive consumer perceptions of the retailer and choices (Lemon and Verhoef, 2016).

2.2.2 Changing preferences induced by e-commerce

Restructuring of the traditional brick-and-mortar retailer landscape through increased internet retail had previously been thought as detrimental to the market share of retailing, leisure and service sectors as consumers progressively sought convenient shopping opportunity. These concerns were linked to the growing trend of physical shopping opportunity being substituted online, with early predictions suggesting high street stores in the UK were estimated to lose 20% of their business to online retailing (Angelides, 1997). This perspective was channelled by a mood of unbridled optimism regarding the internet's potential to reshape the commercial world, which caused a flurry of hyperbole issued by media articles and consultant reports relating to the death of physical retail (Doherty and Ellis-Chadwick, 2010). Concerns centred on opportunities internet retailing provided for manufacturers to target consumers directly, thereby removing the intermediary role of the retailer (Malone et al., 1987). Moreover, commentators envisaged that new players without an established physical presence within shopping destinations could pair commerce software to scheduling (and distribution) capabilities to achieve a lower cost channel structure (Doherty et al., 1999). These changes were perceived to threaten established retailers as restructuring to the distribution channels of consumer products was seen as a circumvention of traditional distributors.

While these fears were, arguably, well-grounded, online retailing has increasingly become linked to complementarity and modification processes that blend traditional retail channels with e-commerce (Poushneh and Vasquez-Parraga, 2017). In the UK's online grocery sector, for example, restructuring is forecasted to increase market size from £11.4Billion in 2018 to £17.3Billion by 2023 (IGD, 2018). These trends have accompanied the proliferation of modern retail channels which have compelled the introduction of new phrases into the retail lexicon, including 'click-and-collect', 'one-click purchasing' and 'freshly clicked' (Jones and Livingstone, 2018). Thus, consumer preferences have shifted toward a complex set of omni-channel interactions with retailers, which posed new logistical, strategic and operational challenges in the management of resources (Hood et al., 2020). These trends began in the UK during the late 1990s when grocery retailers such as Tesco and Asda started to operate an online grocery service using a home delivery network (Clark and Chang, 2014). During this period most retailers adopted store-based online packing and dispatch to command a rapid expansion across catchment areas that were proximate to the end user (Hood et al., 2020). Yet, these typical customer-facing stores were often ill-equipped to accommodate an efficient and effective order assembly, which limited capacity and reduced customer satisfaction (Hübner et al., 2016). Moreover, the challenges associated to this 'last mile' has been argued to account for over 50% of supply chain costs relating to order fulfilment (Aspray et al., 2013). Alternative delivery modes that afford greater logistical efficiency, reduce costs and free consumers from the restrictive time-slots required to accept home-delivery orders compelled retailers to introduce new channels in an attempt to reflect changing consumer behaviours (Hood et al., 2020).

To accommodate these changing consumer preferences, retailers most recently have introduced collection points that allow shoppers to purchase certain products from particular stores without any direct interaction with the store itself (Jones and Livingstone, 2018). Conveniently for consumers, they can then opt to pick up the product at a location that best suits themselves. This strategy allows retailers to uphold and increase a geographical market presence in competitive catchments while also remaining competitive to the offer of online retail (Vyt et al., 2017). In the UK, for example, Jessops, a camera retailer with a strong brand presence, undertook a major restructuring that allowed the business to reduce their physical store network from 187 in 2013 to 51 in 2016 while maintaining healthy revenue streams through online sales (Jones and Livingstone, 2018). This restructuring has often been facilitated by many retailers offering collection from non-store locations like leisure venues and transport interchanges that have proved successful for non-perishable products, but less so for groceries. In November 2013, for example, the UK retailer ASDA explicitly targetted commuters by introducing click and collect services to six London Underground stations that allowed orders before noon to be picked up from the station car park (Odell and Pickford, 2013). To date however, these online initiatives have made virtually no impact on physical floor space occupation, highlighting the difference in consumer behaviours between grocery and non-food sectors, in addition to unique supply-demand interactions (Hood et al., 2020).

Online retail touchpoints

While design elements traditionally pertain to visual features of physical retail spaces, key properties entailing functional and aesthetic elements of online touchpoints (e.g. website, mobile app) are increasingly employed to convert transactions online through schemes such as home delivery and click-and-collect (Roggeveen et al., 2020). Relevant functional elements include layout, navigation, search speed and organisation of retailer websites, while aesthetic properties refer to factors such as emotional appeal and "uniformity of the website's overall graphical look" (Cyr, 2008). Amongst existing research, researchers have found the quality of online merchandise displays are drivers of purchase intention, with high investment sites characterised by white backgrounds, elegant fonts, product videos and enhanced zoom features as increasing preferences and valuations of hedonic options (Schlosser et al., 2006; Roggeveen et al., 2020). In addition, introduction of services that grant authentic situated experiences and trialability through augmented or virtual reality enhance customer value perceptions by allowing product exploration that simulate physical control and environmental embedding (Hilken et al., 2017). Thus, while anecdotal evidence suggests growth in technologies such as online mobile phone use work to reduce point-of-sales purchases, Grewal et al. (2018a) argues restructuring has caused a transcending between distinctions of physical and online retail landscapes, which can lead to increased purchases even when consumers divert from conventional shopping loops.

Changing retail hierarchy

The rise of internet retailing is most associated to opportunities provided by mobile devices for online shopping, such as smartphones and tablets, alongside the penetration of broadband, which by early 2019 had 95% of UK households provisioned by 24Mbps+ connectivity (DCMS, 2013; Singleton et al., 2016). These changes have accompanied an increased consumer demand for e-services, which have carried transformative changes to the hierarchies of physical retail spaces. Growing online consumption has transferred power from retailers to consumers through the provision of opportunity for 24/7 convenience and price comparison alongside a wider potential geographical reach of products to the consumer (Williams, 2009). Evidence suggests the recent expansion in online consumption and digital technology has affected the health of retail centres in complex ways (Singleton et al., 2016). However changes to the geography of physical retail spaces are required to be understood within the wider historical retail context. Beginning in the 1970s, the emergence of motor vehicles transformed consumer shopping behaviour and retail development, resulting in grocery-led flight from traditional town centres (Jones and Livingstone, 2018). Upheaval to the retail hierarchy from online retail, therefore, must be understood as a culmination of structural change across the last fifty years. Similar to the *motor age*, internet retail enhanced opportunities for convenient consumption, which invoked another phase of restructuring for traditional UK high streets (Wrigley and Lambiri, 2014). Amongst this change, the role of retailer size has been argued as an important issue that potentially allows retailers to capitalise upon growing online usage and expand beyond their natural catchments. Large retailers with established brand presence that specialise in comparison and fashion products have generally flourished within this new marketplace characterised by click and collect methods of consumption. Conversely, adoption of new digital technologies have been slower in small- and medium-sized enterprises (SMEs) (Wagner et al., 2003). To account for this, Jones and Livingstone (2018) suggest that while SMEs operating click and deliver are able to expand local catchments, the scope for them to exploit omni-channel online sales is constrained by the limited access to investment capital for the heavy development costs of this structural change.

Changing consumption behaviour enabled by a growth in e-services has, most importantly, heralded investigation into the implications of internet retailing on traditional physical shopping destinations. Singleton et al. (2016) propose a framework of *e-resilience* to quantify the extent of vulnerability for retail hierarchies to the substitution or replacement of physical shopping opportunity, in addition to the blending with traditional retail via complementarity or modification. Their conceptual framework of e-resilience is visualised in Figure 2.3. The central concept of e-resilience estimates the likelihood that existing infrastructure, functions and physical shopping provision of the retail centre can adapt or accommodate the proliferation of internet services. Empirical evidence suggests, for example, the presence of anchor stores and service providers that are difficult to digitise such as leisure venues or outlets that employ experiential marketing to engage with consumers and convert sales outlets are typically associated with lower online substitution rates (Weltevreden, 2007). Conversely, retailers who trade merchandise such as books, music or video games media that is easily digitised are highly receptive to competition from internet retailers (Singleton et al., 2016). This non-uniformity concerning the impact of online shopping across retail types had been documented early in this growing literature (Sinai and Waldfogel, 2004). Yet, the supply and demand factors that influence retail offerings are also contingent on the (geo)demographic characteristics that shape the likely internet engagement behaviour of prospective customers. Below, we briefly consider these factors before concluding our discussion of modern consumption patterns.

Figure 2.3: Singleton et al. (2016)'s proposed conceptual framework of e-resilience.

Geodemographic drivers of shopping behaviour

Geodemographic characteristics of the catchments serviced by retail centres fundamentally drive consumer behaviour and the propensity to partake in internet retail which, by extension, influences the health of established retail hierarchies (Birkin et al., 2002). More concretely, a behavioural component that relates to characteristics of people and their residential location captures engagement with internet retail and so shapes the differential geographies of online shopping (Longley and Singleton, 2009). Across various contexts, preferences for internet retail relate to demography and geographic context. Typically age has played the most important role in driving e-shopping behaviour, with an Acxiom Research Option Poll finding highest engagement from 25–44 year olds, while just 1 in 10 respondents from over 65's reporting consistent e-commerce use (Clarke et al., 2015). Despite this, more recent evidence suggests significant growth in the rate of online purchasing amongst the 65+ group, with 48% purchasing online up from 16% in 2008 (ONS, 2018). These trends suggest that retail centres able to adapt their offerings into the growing virtual marketplace have the potential to boost the vitality and viability of physical shopping destinations. This is because digital technologies grant accessibility to online information of product availability, stores, services and brands prior to visiting, and these convenience factors enhance the overall customer experience enjoyed within the physical shopping destination (Wrigley et al., 2015). Alongside age, gender also acts a dominant factor driving internet shopping engagement. Mortimer et al. (2016) finds higher frequencies of females partake in online shopping compared to men across multiple retail sectors, however the reverse has been shown in the Netherlands (Weltevreden, 2007). In addition, reflections of affluence such as household disposable incomes are key considerations in consumer purchase decision-making. Davies et al. (2019), for example, show increased up-take of store-based click and collect points amongst affluent socio-economic groups, while Weltevreden (2007) finds positive associations between the education of respondents surveyed and the propensity to engage with e-commerce.

Lastly, alongside demographic and socio-economic variation, behavioural patterns also vary according to a spatial component that directly links to the geographies of demand for retail environments. Geographic variations in the ownership of basic digital skills (Helsper and Eynon, 2010), the speed of connection (Singleton et al., 2016), and remoteness of locations (Warren, 2007) are all factors that shape consumer spatial behaviour within the digitally transformed retail landscape. For instance, Hood et al. (2020) argues e-commerce engagement can be differentiated between areas based on rurality, where online shopping propensity increases in places where access to physical stores is lower. Cao et al. (2013) offer an alternative theorisation, claiming e-commerce to be an urban phenomena driven by technology that would slowly diffuse to suburban, exurban and finally rural areas. Irrespective of how e-commerce uptake was theorised to spread, the literature unequivocally argues consumer behaviour *vis-a-vis* internet retail is governed by an interaction between the multidimensional social attributes of particular areas and the underlying geography. Efforts to understand how attributes of residential locations influence consumer behaviour towards e-commerce have commonly relied upon geodemographic classification such as the Internet User Classification (IUC) (Singleton et al., 2016) for differentiating the degree of an area's e-resilience. Clearly, the growth of internet retailing facilitated by consumer preference for digitally-enabled methods of consumption heralded transformative change to the physical hierarchies of shopping spaces. Yet, despite continued uncertainty for traditional high streets, geography remains an important force shaping the vitality of retail centres. These changes are contextualised amongst structural adjustment through multi-channel shopping opportunities and consumer expectations which are increasingly technology-driven.

In conclusion, this discussion has sought to discuss the changing nature and complexities of consumption behaviour in the modern era. In the following section, we critically analyse existing empirical studies that elicit *stated* or *revealed* preferences for particular retail environments.

2.3 Empirical methods for measuring consumer perception

Rushton (1969) famously described consumer spatial behaviour as an outcome of the search among alternative opportunities, with preferences driving the expression of observed consumer tendencies. The actual nature of this behaviour is able to be approximated empirically, and contributions that study possible drivers of retail location preference are typically classified into two streams of research. In this section, we briefly review both before opening a critical discussion contrasting their strengths and limitations.

2.3.1 Stated choice approaches

Stated choice models measure the importance of different objective characteristics from a particular set of choices by introducing specifically-designed questionnaires to samples of individuals (Vyvere, 1994). Typically, attitudinal and perceptual values towards attributes of retail environments are represented by asking respondents to choose an option between a given choice set, or to rank alternatives on a semantic scale to ascertain the level of satisfaction or importance (Morikawa et al., 2002). Choices are often coded into independent binary variables for each attribute of a retail location the analyst is interested in estimating the utility for. Stated choices across all consumers can then be decomposed into partial weights reflecting the importances of each choice alternative within the set (Vyvere, 1994). These approaches are rooted in random utility theory, which postulates that shoppers possess unobservable, latent (and stochastic) preferences when choosing between retail destinations. Assuming the principal of utility maximization, the probability of choosing between a set of alternatives is equal to the probability the utility for a choice option exceeds all other alternatives in the choice set (Timmermans, 2004). When predicting shopping location choice, consumers will evaluate decisions in reference to characteristics described in Section 2.2.1, such as place-based attributes like travel times, tenant mixes and parking facilities (Oppewal et al., 1997). The most widely applied model to estimate stated choices is the multinomial logistic (MNL), where the preference of consumer i towards alternative j is expressed by,

$$P_{ij} = \frac{exp(U_{ij})}{\sum\limits_{q=1}^{J} exp(U_{iq})}, \qquad j = 1 \ldots J \tag{2.2}$$

where P_{ij} represents the probability of consumer i choosing store j from J alternatives, and U_{ij} is the observed utility of store j for consumer i. Moreover, the consumer's utility function U_{ij} for estimation is written as,

$$U_{ij} = \sum_{k=1}^{K} \beta_k x_{kji} + \epsilon_{ij}, \tag{2.3}$$

where x_{kij} represents attributes of store j or socio-economic characteristics of individual i, β_k is the parameter for attribute k and ϵ_{ij} denotes the random, unobserved component of utility (Moore, 1989). Thus, U_{ij} is an additive utility function that reflects a compensatory decision-making process, where lower valuations of particular retail location attributes may be compensated by higher evaluative scores on other, remaining attributes (Timmermans, 2004). Typically for each respondent, a combination of binary variables that reflect the importance of different attributes define a *hypothetical* alternative of choice, which is hypothetical due to its creation within an experimental setting (Moore, 1989).

Amongst this rich empirical literature, studies typically draw small numbers of samples from designated areas before speculating how their findings generalize beyond the study. Below, we introduce the workflow of three studies taken as best practice examples to demonstrate the execution of stated choice approaches. In the first of these, Oppewal and Holyoake (2004) use questionnaire choices derived from 220 undergraduate students at an Australian university (see Figure 2.4) to test a range of hypotheses that assess whether the propensity for multi-purpose shopping behaviour increases when consumers have more information about individual components. An example question asked is, "*Are consumers more likely to buy a separate component (flight) if they know there is a possibility to buy a matching component (accommodation) in a nearby store?*". The authors rely on theory of 'unbundled markets', where components are traded separately and can be self-assembled by consumers (Wilson et al., 1990). Their contribution show how choice experiments dissect the shopping process, demonstrating how consumers equipped with more product information compel them to buy components separately and from different stores.

Options	At current travel agent	At current travel agent	Available at the previous travel agent **20 minutes away**	Keep Searching for other alternatives **(1 other store in your suburb)**
Destination*	Your destination B	Your destination A	Your destination A	
Price	$1100	$600	$400	
What You Are Buying	Flight: Qantas Hotel: Budget Information brochure: YES Complimentary arrival tour	Flight: Garuda	Hotel: Moderate Information Brochure: NO	
Your Choice (Select One Only)	☐ **(Buy Package)**	☐ **(Buy Flight Only)**	If you chose **Flight Only**, would you travel back to the previous agent to buy this accommodation? Yes ☐ No ☐	☐ **(Not buy in this store)**

* Respondents had been instructed that their destination "A" was the destination they had selected earlier as their most preferred new pacific destination; similarly, "B" was their preferred repeat destination (if a respondent had not visited any pacific destination, B was the second preferred new destination).

Figure 2.4: Example choice scenario imagined within stated preference experiments (Oppewal and Holyoake, 2004).

Elsewhere, Arentze et al. (2005) used a stated choice experiment across a sample of 1,704 household telephone surveys in Northern Brabant, the Netherlands, that asked respondents to recall their most recent shopping trip to a retail centre. For instance, participants were required to report whether they had purchased groceries, clothing, or other semi-durable categories across the trip. A nested-logit model was then estimated to reveal evidence for many hypothesised agglomeration effects which were shown to increase the volume of multi-purpose trips. Location, size and perceived purpose-specific utilities of shopping destinations were found to elicit the most significant impact on choice of trip purpose. One caveat of the study is they consider shopping trips in isolation from other out-of-home activities that might interact with the propensity to shop.

Lastly, Severin et al. (2001) used stated choice models to assess the stability of preferences underlying retail-shopping choice over time and space in Edmonton, Canada. Across several years, samples of 740 (1992), 624 (1993) and 476 (1996) consumers were used to assess perceptions of shopping spaces. Survey respondents were required to circle features that applied to particular retail centres – "high quality", "wide selection" and "nice atmosphere", for example. Their research demonstrated stability of underlying drivers of preference over time, which suggested that managerially similar retailing strategies should be expected to yield similar results. One limitation of their approach, however, is that only a small subset of perceptual attributes were common to each of the sample years.

These three studies demonstrate the general workflow of stated choice research, which typically proceeds by hypothesising some phenomena relating to drivers of consumer preference, before planning an experimental design in which to test this. While we describe only three studies for illustration, we note the existence of a rich set of literature within this domain of research (Arnold et al., 1983; Dellaert et al., 1998; Arentze and Timmermans, 2001; Teller et al., 2008; Teller and Elms, 2010, 2012; Zoltan and Masiero, 2012; Badrinarayanan and Becerra, 2019).

2.3.2 Revealed preference approaches

Revealed preference is opposed to *stated* preference, with the former focusing on real-world data and overt human behaviour (Moore, 1989). Rushton (1969) argues consumer spatial behaviour, as any other behaviour, can be determined only through revealed preferences alone. This direction is influenced by the traditional view in econometrics that valid choice data is only derived from actual choices taken place (Morikawa et al., 2002). Thus, revealed preference is an 'ex-post' analysis which estimate choices constrained by factors like budget, time or accessibility, for example. Hedonic price modelling is one valuation method typically used in real estate studies to *reveal* the implicit price of a property's utility-bearing characteristics (Rosen, 1974). These techniques express the value of complex goods like housing as a function of multiple intrinsic and extrinsic characteristics, including structural, locational and environment attributes common to the property.

Alongside residential properties, retail premises can also be understood as a complex good within the framework of hedonic regression. Retail centre design and site location factors are critical to bridging consumers and products, with Salleh and Ruddock (1999) arguing that understanding which attributes influence valuations like rental prices assist retail managers in the provision of better facilities, which drives enjoyable leisure experiences and consumer spending. Most importantly, a retail facility's market position can also be inferred from its paid commercial rent (Hui et al., 2007). This can be decomposed further, as higher commercial rents generally infer more attractive retail spaces which draw consumers from wider geographical catchments due to the gravity of their composite retailers influence (Dennis et al., 2002). Ultimately, holding other things constant, the performance of shopping locations can be reflected by rental value, which further approximates customer perceptions on the brand/marketing positioning of particular consumption spaces (Hui et al., 2007).

The general form of a hedonic regression model for a retail premise can be specified as a linear combination of exogenous factors (Rosiers et al., 2005), expressed simply as:

$$y_i = \alpha + \boldsymbol{x}_i \boldsymbol{\beta} + \epsilon_i, \tag{2.4}$$

where y_i often reflects annualised per-area net rental value for store i, $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients to be estimated, $\boldsymbol{x}_{ik}$ is a $1 \times k$ vector of price determinants for store i, and ϵ_i denotes random unobserved error assumed to be i.i.d. In the literature, typically *non-spatial* and *spatial* determinants of retail premise rents constitute the variables commonly used in the descriptor series, $\boldsymbol{x}$ (Nase et al., 2013). Non-spatial variables include drivers discussed in Section 2.2.1, such as retail image and mix, alongside positive shopping externalities derived from the presence and type of anchor tenants which proxy economic potential and retail gravitation indices (Mejia and Benjamin, 2002; Hui et al., 2007). Non-spatial determinants are so important that Gatzlaff et al. (1994) showed losing an anchor tenant caused 25% reductions in non-anchor tenant rents, illuminating the customer drawing power and externalities retailers derive from positioning alongside key tenants. Aside from anchor tenant presence, non-spatial determinants also include: variety of merchandise; price levels; franchise reputation; customer service quality; and shopping atmosphere and cleanliness (Ibrahim, 2002). Alongside non-spatial drivers, spatial determinants of retail rent prices have also been widely identified in the literature. These include site location and structural design characteristics, which consist of physical attributes such as: retail centre shape (U-shape or L-shape, for example); density of nearby shops; store positioning relative to high footfall locations; leaseable gross floor area; number of transportation links nearby; and the amount of urban greenery present (Rosiers et al., 2005; Borst et al., 2008; Teller and Elms, 2010).

While stated choice research belongs to a rich history, hedonic studies that elicit revealed preferences through retail property rents are still embryonic due to confidentiality concerns associated with retail transactions (Rosiers et al., 2005). Nonetheless, below we introduce three studies observed as best practice examples for eliciting revealed preference through retail hedonic approaches. Our first example, Hui et al. (2007), explores the relationship between market positioning and retail rents across 151 stores in Hong Kong. A regression analysis is deployed to learn associations between per-area net rental value and attributes describing physical characteristics, location factors and market position (district centre or high street shop, for example)[2]. The study shows district centres as commanding the highest average rental values, presumably owing to positive shopping externalities derived from many stores at a single location, which reduces transport and product search costs. A major caveat of their findings, however, is the small sample size, which limits the generalizability of this research.

Cat.		Symbol	Unit of measure	Descriptions
X	Rent	*R*	HK$ per m^2 p.a.	Net rental income per unit area of the shop per year
	Age	AGE	Years	Age of the shopping mall
	Total floor area	GFA	m^2	Gross floor area of the shopping mall
	No. of shops	NOS	Each	Number of shops in the shopping mall
	Occupancy rate	OCR	%	OCR = (Occupied shop/No. of shops) × 100 percent
	Efficiency ratio	EFR	%	EFR = (Lettable floor area/GFA) × 100 percent
P	District center	DC	Dummy	DC = 1 if the mall is a district center, zero otherwise
	Estate center	EC	Dummy	EC = 1 if the mall is an estate center, zero otherwise
	Local center	LC	Dummy	EC = 1 if the mall is a local center, zero otherwise
	Shop	SH	Dummy	EC = 1 if the mall is a shop on street, zero otherwise
L	Tseung Kwan O	TKO	Dummy	TKO = 1 if the mall is located in Tseung Kwan O, zero otherwise
	Tuen Mun	TMN	Dummy	TMN = 1 if the mall is located in Tuen Mun
	Tai Po	TPO	Dummy	TPO = 1 if the mall is located in Tai Po
	Yuen Long	YLG	Dummy	YLG = 1 if the mall is located in Yuen Long
	Kowloon East	KLE	Dummy	KLE = 1 if the mall is located in Kowloon East
	Kowloon West	KLW	Dummy	KLW = 1 if the mall is located in Kowloon West
	Kowloon Central	KLC	Dummy	KLC = 1 if the mall is located in Kowloon Central
	Kwai Chung	KWC	Dummy	KWC = 1 if the mall is located in Kwai Chung
	Ma On Shan	MOS	Dummy	MOS = 1 if the mall is located in Ma On Shan
	Hong Kong Island	HKI	Dummy	HKI = 1 if the mall is located in Hong Kong Island

Table II.
Definitions of variables

Figure 2.5: Example set of variables used in retail hedonic studies (Hui et al., 2007).

The second example, Nase et al. (2013), focuses particular attention on the relationship between urban design quality and the real estate values of 301 retail premises in Belfast, Northern Ireland. A regression-based approach was estimated utilising a range of variables designed using quantitative and qualitative approaches to strengthen the research value. The authors find aspects of design including store frontage continuity and variety, connectivity and building material quality all increased valuations of rental values, while also emphasising the high impact of location and tenant mix characteristics. Ultimately, their findings *reveal* which aspects of quality design are highly valued by retail tenants and, by extension, elicit favourable consumer perceptions. Unfortunately, the study suffers

[2]Complete set of variables displayed in Figure 2.5.

similar limitations relating to the restricted availability of data describing large numbers of retail units.

Most recently, Koster et al. (2019) use instrumental variables (IV) estimation – a standard econometric technique – to recover causal relationships between shopping externalities like footfall and numbers of shops and rental values across 4,738 retail units in Amsterdam, the Netherlands. The authors control for potential endogeneity by employing an instrument that uses exact cinema locations from 1930, which has strong autocorrelation with footfall and the number of shops in the vicinity. By removing sources of endogeneity, the authors arrive at causal claims that demonstrate shop rents as positively dependent on high footfall locations, with a store's marginal willingness to pay for a passing pedestrian amounting to around 0.009 euros. However, one limitation of their study is the highly peculiar consumption landscape of Amsterdam, which potentially limits generalizability of their findings.

As this brief review confirms, hedonic investigation of retail properties are still embryonic. Sector peculiarities mean the existing retail hedonic literature is often hampered with small quantities of rental data, with this fractional number of studies focusing most often on planned retail agglomerations such as shopping malls (Benjamin et al., 1990; Sirmans and Guidry, 1993; Mejia and Benjamin, 2002; Rosiers et al., 2005), but also unplanned clusters like neighbourhood centres (Hardin and Wolverton, 2001; Hui et al., 2007; Nase et al., 2013; Koster et al., 2019). Having outlined the workflow of *stated* and *revealed* preference research, in the following section we critically discuss the applicability of both approaches to research questions in retail geography.

2.3.3 Strengths, limitations and contrasts

Consumer spatial behaviour, in general, can be viewed in two opposing fashions. Eliciting preference through direct questioning reflects an outcome of choices reflecting motivations and values of individuals, while observation of overt behaviours reflects the constraints of environmental and personal circumstances (Pahl, 1970). In both cases, each statistical unit of observation reflects either the signalled intention of choice or marker of actual behaviour, such as willingness to pay (Vyvere, 1994). While stated preference techniques have been used extensively by market researchers, revealed preference models are the domain of econometricians, who argue valid choice is observable only from actual choices being made (Morikawa et al., 2002).

The case for stated preference research

Early growth in popularity of stated preference modelling stemmed from dissatisfaction of the interpretive performance of aggregate econometric models (Louvière and Timmermans, 1990). Significant advances in experimental design allowed retail analysts to study hypotheses through stated choice experiments that were rigorously controlled (Vyvere, 1994). Market researchers argue that using observational studies to test utility functions are often constrained by factors imposed in real markets, such as statistical confounding from inter-variable correlation and endogeneity from omitted, unobservable characteristics. Unless revealed preference data exhibit well-behaved properties, the estimates of utilities consumer derive from particular retail characteristics are likely to be biased. This is because of the boundless constraints that limit the likelihood revealed preferences translate to overt behaviour, which raises the ontological question of whether one is observing the reality of a desired outcome. As Hwang and Albrecht (1987) remark in context with residential housing, “there is a persistent discrepancy between residential preferences and actual moving behaviour", which applies equally to retail premises. Preferences are ruptured by high costs of purchasing (or renting) property, and the immobility of housing goods (Vyvere, 1994). For example, a retail manager might be forced to lease a shopping unit that costs the same as a unit at a destination that naturally attracts more passers-by as a result of limited supply. In this case, higher rental values do not necessarily infer more attractive retail spaces that are able to draw high volumes of footfall, as is suggested in the retail hedonic literature. Anderson (1971) goes further to argue revealed preferences reflect a danger to research, as observed behaviour might be misinterpreted to reflect what people *choose* rather than *are forced to do*. Ultimately, the authors above stress stated preference data should not be dismissed too readily when compared to revealed preferences, as care must be taken to ensure assumptions in the latter are not too strong for a particular application.

Another advantage of stated preference approaches are that the hypothetical set of choices determined within an experiment are independent of choice context, and can be designed to cover a range of attributes for inclusion to the modelling procedure (Moore, 1989). Respondents can be questioned on preferences for stores with, for example, diverse variety of merchandise, presence of leisure facilities such as bars and restaurants, or availability of free parking, even though the shopping destination may not contain such options at the time of study. There is no free lunch, however, as Louviere et al. (2000) warns of the importance for new, proposed alternatives to not be too dissimilar from existing choices. Reasonable familiarity with the range of alternatives presented within the stated preference experiment is necessary, as little familiarity means respondents formulate no preferences or discriminating rules when evaluating attribute utility when choosing between stores (Moore, 1989).

Finally, stated preference approaches carry several empirically-driven advantages when compared to revealed preference methods. As variation in attributes described within the chosen set of survey questions are entirely specified by the analyst, stated preference analysis allows one to control correlations between variables (Vyvere, 1994). This is particularly useful for unpacking the multidimensional nature of how particular variables drive retail location choice. Conversely, in revealed preference research, distinguishing independent effects from the explanatory variables are likely to be tempered by inter-correlations that exist between real-world variables (Moore, 1989). For example, retail environments with a high density of stores will generally also have a large number of parking spaces. Thus, unless orthogonality is observed among these variables, measuring their separate contribution to the consumer's utility function will be biased due to correlation cancelling out their effects in the model (Moore, 1989)[3]. Lastly, revealed preference models are also constrained to systems and choice processes based on the observed data. This means they cannot make inferences or predictions of the likely effects of radical new alternatives and non-existing attributes, simply because revealed preference data accommodate no information about their effects on present experiences (Vyvere, 1994; Morikawa et al., 2002).

The case for revealed preference research

Protagonists of revealed preference research often cite the philosophical works of Bertrand Russell, who evinced that dependable preference information is yielded from what their choices reveal them to prefer, rather than what individuals proclaim to (Pirie, 1976). Economists argue preferences should satisfy consistency with what psychologists call *construct validity* (Schlaepfer and Fischhoff, 2010). That is, the appropriateness of inferences

[3]Stated preference models are, however, not exempt from these challenges. A significant criticism against MNL models used in choice analysis is that the utility contribution of a choice alternative is independent from the set (and attributes) of other alternatives in the choice set – this is known as the Independence from Irrelevant Alternatives (IIA) property. In reality, this assumption is too strong as particular alternatives will exhibit a high degree of similarity (Timmermans, 2001; Severin et al., 2001). To circumvent this limitation, extensions of the basic MNL model have been proposed to relax the IIA assumption. The nested logit model, for example, mitigates the IIA-property by assuming correlated choice alternatives group into the same nest (Arentze et al., 2005).

made on the basis of observations and insensitivity to irrelevant information. A well-known violation of these conditions arise from *anchoring biases*, whereby responses are guided by values within questions that respondents use as mental reference points when making decisions (McFadden, 2001). An example hypothetical question that inserts an anchoring effect would be: "How much would you pay for parking at this retail location? [£2, £4, £6]". In this example, the anchor encodes information as to what the researcher believes the value of parking to be, rather than this information being inferred empirically from the sample. Additionally, stated preference approaches often expose respondents to retail location evaluation tasks in less life-like ways – telephone surveys or interviews undertaken away from retail locations, for example. Tasks in these circumstances require respondents to invoke strong imaginary skills when simulating a realistic shopping situation in their mind (Teller and Reutterer, 2008). In doing so, these study designs ignore the situational and shopping context-specific factors that consumers evaluate under real shopping situations.

One direct advantage of revealed preference approaches are that they portray retail systems in a state of market equilibrium with fixed constraints (Abdullah et al., 2011). There are considerable motivations explaining why one might expect preferences elicited through stated preference questioning to differ from realistic experiences. For example, consumers may fail to adequately account for budgetary constraints and feasible substitute products and retail locations when responding to hypothetical scenarios of choice presented within the survey (Azevedo et al., 2003). Time and financial resources consumers perceive as available for allocation toward particular locations and products may be inflated by poor survey design (Bateman et al., 2002). Low scenario credibility ensues when consumers have uncertainty over valuations of attributes under unfamiliar price and quality settings. Thus, the preference structure governing revealed preferences might be completely independent of the structure governing stated preferences (McConnell and Strand, 1981). When true, this infers information derived from stated preferences is arbitrary, leading to unreliable estimates of consumer utility for particular attributes. These issues are corrected under revealed preference data, because they are constructed by users in a determined context of constraints (dell'Olio et al., 2018).

Another advantage of revealed preference methods is they arrive more easily at external validity, which is definable as the ability of a decompositional model to explain real-world behaviour. As Rushton (1969) explains, revealed space preference is the organising principle of consumer spatial behaviour. Rankings and orderings of hypothetical sets of retail opportunities based on common observed characteristics – such as whether retail tenants pay more for shopping destinations that attract high footfall – are *likely* to be place independent, and *unlikely* to vary between places unless major cultural boundaries are crossed. In stated preference approaches, on the other hand, it remains unclear the extent that findings are useful in understanding real-world processes and decisions outside the sample (Horowitz and Louviere, 1995). This problem relates to the geographic transferability of models calibrated in one location being unable to accurately predict choices in another, which occurs due to differences in socio-economic constraints that shape consumer behavioural tendencies across different market segments and geographic regions (Vyvere, 1994). This issue is limited in revealed preference approaches because they accommodate such constraints as fixed. One way to test external validity of stated preference approaches are to apply 'before/after' studies, where stated preferences are confirmed by subsequent choice behaviour (Wardman, 1988). Unfortunately, however, this adds an additional layer of complexity to the study design.

An ongoing debate among researchers implies no single answer to the question of which, *stated* or *revealed* preference methods, are the most theoretically and empirically valid for understanding consumer perceptions of retail environments. A sensible resolution would be to realise there may be answers to where (or when) the approaches are more suitable (Shogren, 2006). Across this section, we have reviewed the two principal streams of research that elicit consumer preferential behaviour. In both cases, each stream typically use small samples of data that exhibit well-behaved tendencies, where data points are neatly delimited into rows and columns of a spreadsheet. The research questions explored within this thesis use data that does not observe these well-behaved tendencies, and requires new epistemological frameworks of thinking, which we explore in the next section.

2.4 Machine learning applications in urban science

In recent years, the ability to quantify urban phenomena in contexts where data were traditionally scarce has been transformed through access to a 'data landscape' that increasingly traces aspects of human behaviour (Arribas-Bel, 2014). A growth in the amount and diversity of structured and unstructured data sources available to urban researchers has enabled the testing of hypotheses concerning emerging patterns that were hitherto unthinkable. Previous data-informed urbanism premised ideas and theory on data such as censuses, household surveys and commissioned interviews that were typically sampled at highly coarse temporal granularities, offering occasional snapshots into urban spaces (Kitchin, 2016). Increasingly however, new forms of urban big data that are exhaustive in scope are able to augment these traditional datasets. This transformation has been enabled by the roll-out of digital technologies and Internet of Things (IoT) platforms that collate huge volumes of data from various sources including: sensors; smart systems; online communications; audio and video; search queries; and media files (Rathore et al., 2018). These technologies embed within the urban fabric and continuously send data to control and management systems. In context of urban research, sensors, for example, can be used to monitor MAC addresses to infer mobility of pedestrians across space, which can even be contextualised alongside data that are sampled less periodically like public census records. To illustrate the scale of storage requirements as a by-product of this embedding process, Arribas-Bel and Reades (2018) show the entirety of US Census data is a mere 6.3 GB when compressed, while self-driving vehicles are estimated to produce over 4 TB of data every

90 minutes. The consequence of this emerging *data deluge* concerns the subtle replacement of a data-informed to a data-driven urban science. Academic researchers aided by data analytics software are able to process, analyse and visualise this vast deluge of data for problem solving at fine spatio-temporal resolutions, while analysts from the public domain may use big data systems for a highly responsive urban governance by influencing or controlling how city systems perform and respond (Kitchin, 2016).

A principal driver of this recent change lies in the growing availability of cheap computational hardware to accommodate the ingestion, warehousing and processing of urban big data. The declining size and cost of computer software and hardware platforms enables the possibility of building networked devices that innovate the collection of geographical data, but also the dispersion of this data through application programming interfaces (API) (Arribas-Bel and Reades, 2018). Google Street View (GSV) is one such API that obtains 360° panorama images of streetscapes from recordings made by the roof-mounted camera of cars that tour road networks of cities and rural areas worldwide (Anguelov et al., 2010). Crucially, each unique panoid is accompanied by metadata that describes the latitude, longitude and time stamp of a location when an image is recorded (Naik et al., 2017). As we discuss later in Section 2.4.1, the cumulative ability of systems like GSV to record aspects of urban environments with unprecedented coverage is instrumental to unlocking insights from social and physical environments.

But this changing data landscape, alongside the co-evolution of hardware, requires the development of new data analytics that utilise machine learning to process these enormous datasets (Kitchin, 2016). Unlike observations derived from designed experiments, new forms of data often arise without predefined purpose, meaning there is often great difficulty in extracting useful features (Li et al., 2013). This data requires new techniques to extract signal from noise that remain in infancy when compared to traditional statistical methods designed for *data-scarce science* of smaller, clean samples with known properties (Silver, 2012; Batty, 2012). As a response, significant innovation in the fields of data mining, pattern recognition and statistical analysis have recently been made within the computer science community, which have since infiltrated the social sciences, in particular the sub-fields of quantitative geography and urban modelling (Kandt and Batty, 2020). Yet this initial introduction of new methods was roundly criticised within the social sciences at first owing to the brash framing that urban issues could be solvable through technical solutions alone (Kitchin, 2016). Shunning of past scientific practice through a 'data-driven epistemology' was argued as undesirable because the predictive mechanisms of these new approaches are often difficult to interpret and devoid of geographical and social context (O'Neil, 2016). In scientific enquiry, while commercial systems may only require an accurate prediction, social scientists are more concerned with processes that drive outcomes, which has driven this understandable backlash (Singleton and Arribas-Bel, 2019). More concretely, this alleged reductionist thinking is seen to collapse multidimensional social structures to rows in a database, which explicitly ignore the complex aspects of human life that shape urban relations (Kitchin, 2016). Thus, a creation of hypotheses and knowledge 'born from data' as opposed to 'born from theory' is a grave concern.

Allaying these concerns, Singleton and Arribas-Bel (2019) argue for the mediatory role of geography in advancing a new data-driven epistemology that more closely aligns with scientific enquiry, alongside core critical (and ethical) principles. In the remainder of this section, we critically examine the extent to which studies that use methods similar to those applied within this dissertation contextualise processes and findings amongst urban theory. This critical discussion is paramount, as any machine learning innovations in retail geography have to stem from research activities that occur more generally within urban environments.

2.4.1 Urban science studies

The abundance of sensors that receive streams of data like street-level imagery have enabled the possibility of large-scale understanding on how individuals perceive urban environments. For example, research questions that measure or detect shifts in urban perception towards land use across an unprecedented geographical reach have become enabled by this increased digitisation. Alongside the data deluge, computer vision tasks have too become highly popular in urban sciences for automating tasks of the human visual system. Computer vision models are machine learning approaches trained to learn cognitive understandings of visual features from imagery based upon a sequential sample presented to the model (LeCun et al., 2015), and are an example of one method used among the empirical chapters. Throughout this section, we critically review data sources and studies within urban science that apply approaches mirroring those used by this dissertation.

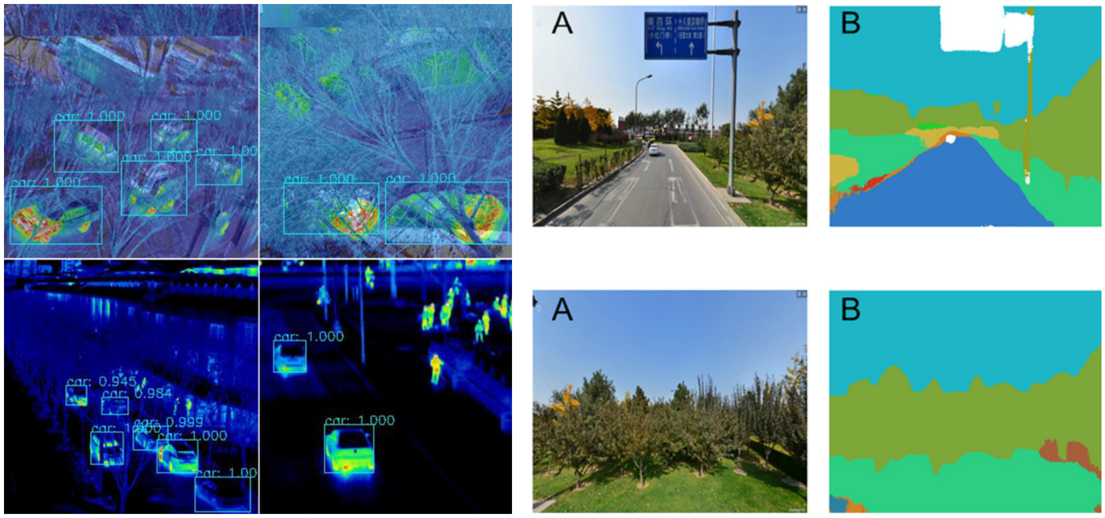
We first turn attention to the conditions that nurtured our ability to understand urban environments at a fine spatial resolution. As shown by Arribas-Bel (2014), the predominant mechanism for this general change in the data landscape resulted from a three-way movement of citizens producing data, businesses moving online, and governments at every level opening their data infrastructure to the cyberspace. For the latter, a release of open data from public organisations have been motivated by the four strategic drivers of transparency, economic value, service improvement and employment generation (Arribas-Bel, 2014). Ancillary to this, data providers such as the Consumer Data Research Centre (CDRC) have pre-processed a number of government sources to increase their usability for the end-user (CDRC, 2020). Elsewhere, in the case of businesses, growing internet popularisation facilitated the digital translation of business activity online. In some cases, businesses have released data in either a reduced volume capacity (Twitter, 2020; Foursquare, 2020) or via aggregate quantities (Google, 2020). In context with this dissertation, the data provider's main avenue for revealing insight to their clients is through virtual dashboards, which require the acquisition of relevant data for building metrics that populate the visualisations (LDC, 2020b). Often a by-product of this digitisation is the construction of datasets that offer unprecedented insight into urban environments that companies offer to academic partners in exchange for knowledge spillover and collaboration. This aspect of the data landscape is the avenue followed by this dissertation, which we discuss in far greater detail in Section 2.5.

More specifically, recent advances in the digitisation of materials required to audit places have enabled large-scale studies to increase the ease of measuring the physical city, and permit learnings of how behaviour might be influenced by urban environments at micro-level. Photographs, surveys and field work were the traditional domain of designers and planners for collecting empirical data for many years, but recent avenues like Google Street View (GSV) have offered more flexibility to automate understandings of urban environments (Yin and Wang, 2016). As argued by Yin et al. (2015) and Ploeger et al. (2016), automation of street-level GSV images through machine learning practically eliminates expensive field logistics, increases coverage beyond traditionally small sample areas, and has been shown to increase time-savings by 50% when collecting data. But while these sources reduce time requirements of direct auditing, critics argue street view data cannot capture every feature that reconstructs the spectrum of perceptual qualities experienced within urban environments (Rundle et al., 2011). Properties intrinsic to human perception like variation in sound and smell, markers of physical disorder like litter or graffiti, and indicators like traffic speeds and perceived safety are difficult, if impossible, to evaluate with static video images (Salesses et al., 2013; Nguyen et al., 2019). Nonetheless, several studies have found agreeability between the results of observational field studies and street-level image audits (Kelly et al., 2013; Hara et al., 2013). While these approaches compare fairly objective measures such as building heights, the recent plethora of research in this area is instructive of how resources like GSV are enabling large-scale understandings of urban environments.

A rich literature exists that pairs computer vision with street-level imagery, with the best examples providing supportive evidence towards classical theories of urban planning. Typically studies 'train' a computer vision model to automatically classify and label images based on discriminating features encoded by the image (Ibrahim et al., 2020). Naik et al. (2017), for example, investigate theories of change in the physical appearances of neighbourhoods for time-series street-level imagery. They train a computer vision model based on data collected from human-derived ratings for the perception of a street's safety. This is achieved using a web-based platform that records user responses to questions that ask, from a pair of images, "Which place looks safer?" The dataset was constructed by the Massachusetts Institute of Technology (MIT) in 2011 as the Place Pulse Project (Salesses et al., 2013), and has since been used extensively across the literature (Dhar et al., 2011; Dubey et al., 2016; Zhang et al., 2018). This crowd-sourced dataset contains 1.17 million pairwise comparisons across 110,988 images and describes perception along a number of domains, including liveliness, affluence and beauty for 56 international cities. Naik et al. (2017) generate 'Streetscores' using the Microsoft Trueskill (Herbrich et al., 2007) rating system for street-view images between 2007 and 2014, and find evidence for the tipping theory of Schelling (1969) that suggests neighbourhoods in poor physical condition progressively worsen over time, as nicer areas get better. While the study is instructive of how quantitative approaches can reveal temporal variation in neighbourhood conditions, GSV images remain proprietary data which offers a pertinent reminder they are collected by a for-profit organisation. Locations that are updated more frequently are driven by searches that drive web traffic to the Street View website, meaning observing change for suburban communities outside urban centres may be difficult because of lower update rates (Yin et al., 2015).

Elsewhere, Salesses et al. (2013) measure the safety, class and uniqueness across four international cities, finding the range of perceptions to be wider, more contrasting and unequal in New York and Boston than Linz and Salzburg. The authors form direct connection to Jane Jacob's Broken Windows Theory that suggests visible evidence of physical disorder, like broken window panes, progressively induce other forms of disorder leading to negative social outcomes and increased crime levels (Jacobs, 1961). Overall, they show evidence to support this theory, and find significant correlation between safety perceptions and homicide rates in New York city. Research analogous to Salesses et al. (2013) follow a similar direction and demonstrate that capturing detailed information describing evaluative dimensions of street-level imagery can be used to characterise urban phenomena such as: walkability (Blecic et al., 2018); gentrification (Ilic et al., 2019); and building quality (Liu et al., 2017).

Another application of computer vision methods relates to the extraction of particular visual elements from urban environments. This direction is influenced by Kevin Lynch's idea of the city's mental map, that particular visual features of urban environments distinguish one locale from another (Lynch, 1960). These studies rely on computer vision models that use object detection and semantic parsing (see Figure 2.6) to automatically identify visual elements such as windows, cars, or people that would otherwise be inferred from manual photograph review or field surveys. Quercia et al. (2014), for instance, find that "visual words" from happy urban scenes are associated with public gardens, red bricks, residential trees and Victoria houses, while unhappy scenes contain more highway road signs and council housing. Gebru et al. (2017) use object detection to identify the make, model and year of all motor vehicles in 50 million images of street scenes in 200 United States cities. They find powerful associations between car make with characteristics such as income, education and voting behaviour, with cities containing a higher number of Sedans than pick-up trucks more likely to vote Democrat than republican. Elsewhere, Helbich et al. (2019) apply semantic segmentation on the pixel space of street-level imagery to identify green (e.g. lawns, street trees) and blue (e.g. rivers, lakes) spaces. Overall they find inverse associations between metrics describing exposure to green and blue space with geriatric depression in Beijing, China. Several other studies highlight the applicability of



(a) (b)

Figure 2.6: Demonstration of object detection (a) and semantic parsing (b) of urban environments (Helbich et al., 2019; Yang et al., 2019).

object detection and semantic parsing for automating neighbourhood quality assessment (Nguyen et al., 2018), learning associations between perception and visual elements (Zhang et al., 2018), and the detection of cars from unmanned aerial vehicles (Yang et al., 2019).

Despite these successes, drawbacks have been identified that argue associations between visual elements and human behaviour typically reflect a predictable response regarding which urban features evoke particular reactions. For example, a removal of man-made objects (which Quercia et al. (2014) show to be associated with ugly scenes) for increasing the beauty of urban environments represent a 'democratic' solution, but possibly not one that aligns to modernist or forward-thinking approaches to planning. If planners are to use study findings to implement one-size-fits-all policies, they face pigeon-holing certain types of buildings, greenspaces or architectural styles based on highly context-specific outcomes, rather than considering the unique variation within their own cities (Quercia et al., 2014; Bader et al., 2017).

Despite this critique, understanding the processes that drive individual and collective outcomes require an eclectic range of neighbourhood conditions (Small and Feldman, 2012). The narrow geographic scope of studies relying on the dispatch of surveying teams limit the generalizability of findings as physical attributes of neighbourhoods have been shown to contrast highly across space (Salesses et al., 2013; Bader et al., 2017). This has led to arguments that only research automated by Street View imagery and computer vision can feasibly record urban appearance at city-wide scales (Naik et al., 2017). Yet, automation invokes a trade-off between spatial resolution and the quality of visual audit. Unfortunately, a balance between high quality auditing and sufficient sample sizes within (and between) neighbourhoods is a requirement to be mediated (Yin et al., 2015). Training a computer vision model requires collecting massive volumes of human-labelled data, and often large, non-expert workforces are utilised to build datasets such as Place Pulse (Salesses et al., 2013). From an economic efficiency standpoint, researchers seek a swift image labelling process for the lowest price, while enabling workers to easily (and accurately) provide annotation. Crowd-sourcing platforms such as Amazon Mechanical Turk (AMT) enable this by allowing researchers to hire pools of online workers to complete annotation tasks submitted to the platform (Sorokin and Forsyth, 2008). Yet, while platforms such as AMT offer resource-efficient opportunity, critics argue urban experiences are highly socially constructed, which introduces human biases into the workflow (Quercia et al., 2014). The user action in response to the question, "Does this place look wealthy? [YES/NO]", for example, is likely to be influenced by shared priors drawn from cultural tastes and socio-economic factors unique to the labeller. Individuals engage with urban environments in different ways, meaning visual characteristics of places may not necessarily convey the same meaning between various socio-demographic groups, which introduces biases into the annotation process if a representative sample of labellers are not pre-selected.

In summary, while data- and methods-driven innovations face considerable challenges, their applications demonstrate substantial promise in building understandings of urban systems related to the built environment. As retail is intrinsically tied to the urban fabric, it is surprising these same innovations have yet to take hold within retail geography. In response, our final section introduces motivations for the development of cross-pollination between the fields of retail geography, urban analytics and machine learning.

2.5 Machine learning applications in retail geography

As a research domain, retail geography possesses many appealing characteristics that are opportune for integration with the ongoing big data revolution. Foremost amongst these, the size of the retailing industry as a sector essential to the livelihoods of families mean it raises potential for research to benefit many stakeholders, such as consumers, manufacturers, retailers and policy makers (Dekimpe, 2019). Moreover, given the sector's multi-faceted nature and constant flux, some argue the sector faces more complex challenges than those elsewhere (Reinartz et al., 2011), which, from a research perspective, invites a more interesting line of questioning. Throughout this section, we document the circumstances that nurture the grounding for which retail geography can capitalise upon opportunities provided by the changing data landscape. Following this, we discuss potential methodological spillovers from the urban sciences that have the capacity to change how hypotheses in retail geography are approached.

2.5.1 Retail sector and the changing data landscape

The retail sector is inundated by data describing characteristics ranging from consumer transactions to the estimated spheres of influence exerted by retail centres (Grewal et al., 2017). A big data industry by definition, retailers like Walmart manage over 11,000 stores across 25 countries internationally, serving over 35 million consumers day-to-day, who purchase an average of 140,000 products across its supercentres (Dekimpe, 2019). According to some estimates, these volumes resolve to around 2.5 petabytes of information each hour, reflecting attributes such as micro-transactions, customer engagement, loyalty expenses and location-based factors (Bradlow et al., 2017). This sheer volume of data collected by retailers is often enriched by linkage with information relating to items such as: inventory statuses across the supply chain (Ulrich et al., 2019); sensors that automate footfall collection (Mumford et al., 2020); location-specific weather data (Verstraete et al., 2019; Badorf and Hoberg, 2020); and indicators derived from social media (Kim and Ko, 2012; Lu and Miller, 2019). For accurate predictive analytics, retailers rely upon timely responses to data signals so adjustment to stock-keeping units are optimised in near real-time (Dekimpe, 2019). Accuracy in these adjustments are critical to avoiding scenarios such as stock-outs, pricing errors or suboptimal recommendations that drive customer attrition (Bradlow et al., 2017).

Additional sources of data capture identifiable consumer- and household-information such as geodemographics, purchase history, and customer locations. This location-based data has typically been generated by growing consumer preferences to use devices like smartphones for shopping, with Ghose and Han (2014) finding that in-app purchase options tend to increase overall app demand. Data collected by smartphones is often mined of geo-location, navigation and usage data from the retailer's shopping app, which provides contextual information of real-time locations that can be exploited to target marketing messages on deals and promotions (Bradlow et al., 2017). In the extreme case, targetting condenses customer segment sizes to size one – also known as personalisation. This granularity of targeting has resulted through use of technologies such as browser tracking cookies that enable retailers to target advertisement at consumers, and even re-target those who exit the virtual store (Kannan and Li, 2017). All things equal, personalisation is generally desired by consumers, as reshaping recommendations to the consumer's unique needs and preferences drives high quality service experiences. Yet, communicating individual preferences requires surrendering personally-identifiable information, which induces a trade-off between personalisation and privacy protection, with strong, recent laws such as the General Data Protection Act greatly restricting the storage of personal information (Rust, 2019). Arguments in this debate proposition that just because certain data are accessible does not qualify them for ethical use, with reminders that customer trust is broken far easier than it is gained (Boyd and Crawford, 2012).

Another distinct source draws upon habitual patterns and consumer behaviours that individuals reveal at the point of sale or when customers patronize a particular shopping destination. Online, examples include eye-tracking technologies for assessing which stimuli draw visual attention from pre-shopping to the point of sale (Huddleston et al., 2018). Offline, the dynamic paths consumers take and the footfall retail spaces generate have been collected as key indicators of town centre vitality and viability (Dolega et al., 2019). Traditionally, footfall (or pedestrian) counts were used by field teams to link patterns of visitor behaviour to the attractiveness of retail locations, and their ability to satisfy catchment needs (Mumford et al., 2020). More recently, WiFi sensors that ping cell phone devices have enabled measurement of actual day-to-day visitor behaviour across wide temporal windows, offering richer interpretations of when retail spaces are used and accessed (Murcio et al., 2018).

Finally, data produced as by-products of efforts to quantify the vitality of shopping environments have emerged *accidentally* which, while intended for different purposes, are useful to retail analysts for testing hypotheses in this area (Arribas-Bel, 2014). The Local Data Company (LDC), for example, hires ground teams of auditors who physically examine every retail location across the UK to construct highly precise and dynamic readings of market trends (LDC, 2020a). Through strategic consulting, the LDC extrapolate present and future trends to their clients, enabling them to make data-driven decisions tailored to their business. Arriving at these insights requires linkage of data collected by the field teams, which include characteristics such as floor space of the premise, full address, type of business, property tax paid for the space, vacancy status, and storefront images. While the LDC create business value from the amalgamation of this data, independently the variables provide potential gold mines of coverage describing retail environments for researchers. Knowledge exchanges between private companies and academic institutions facilitated by businesses opening their data create opportunities to test hypotheses that would otherwise remain impractical. Although this data, like others, is not free of challenge.

Opportunities and challenges

Despite the volume, variety, velocity, and veracity of data in the retail sector, there exists a considerable discrepancy between its potential business value and the ease of extracting those benefits (Dekimpe, 2019). While big (and new) data offer retailers opportunities to create value, guidance of "where to look" amongst the data, alongside clearly-defined hypotheses and outcomes informed by theory, are required to navigate this landscape (Bradlow et al., 2017). Otherwise, focused attention on big data and analytics can overwhelm interpretive efforts to align results with theory, which is crucial to unpacking insight from the generative processes that underlie a researcher's findings (Lehmann, 2020). For example, when Walmart opened their *Express* stores in remote locations with small populations, they maintained identical predictive analytics for setting prices to their other stores, which caused profitability to plummet, with the project closing after four years (Bell, 2014). In this instance, reference to local demography and theories of service uptake in rural environments would have allowed the retailer to distinguish from competitors using local store formats or pricing strategies compliant with area demographics (Kumar et al., 2017).

At its core, theory delineates the story that drives particular outcomes, and is agnostic to the data collection and analysis stages of research (Lehmann, 2020). The importance of theory helps to account for unobservable, latent forces that drive outcomes and to identify spurious trends in the data. Endogeneity relating to variables omitted from the analysis is often tied to biases in model estimates that deceive decision-making efforts. Bradlow et al. (2017) use the example of optimizing advertisement promotions, stating that while the success of physical advertising is contingent on being viewed, viewership is rarely known because of practical challenges in recording this data. But, with knowledge of what data are required, researchers can opt for exploration of technologies like eye-tracking to record consumer reception towards particular adverts, and arrive at more informed decisions by doing so. In this instance, understanding the underpinnings of what data (eye-tracking) was required to assess the key performance indicator (viewership) built insight towards an outcome that optimized the advert display (Bradlow et al., 2017).

More generally, while big data offer potential opportunities for learning new insight, Dekimpe (2019) advises retailers to look beyond the hype, as biases manifest in the reporting of successful machine learning deployment create false impressions of success frequency. Weinberg et al. (2013), for example, argue successful ventures are often perpetuated by technology firms in order to further the adoption of their data-driven solutions by major retailers. Moreover, retail analysts should not blindly accept that databases are free of error and biases, which are magnified when linked with additional datasets of similar properties. As a corollary, empirical studies that examine small samples should not be dismissed too readily also. Small data encourages the necessity of theory and speculation of how findings drawn from small samples generalize beyond the study (Lehmann, 2020). Thus, while data-driven solutions are useful for fine-tuning recurring problems like recommendation offers to loyalty-card holders or stock allocation, they should not necessarily replace small sample studies unless underlined by convincing theoretical motivations.

2.5.2 Opportunities for measuring consumer perception

While retail-based studies are typically encouraged to possess a theoretical basis, not least because of the expectations set by the peer review process, Lehmann (2020) argues such demands potentially occlude "innovative explorations". Absence of a clear predecessor to an analyst's work mean studies often have to pigeon-hole results into fitting a particular theory, which encourages the recycling of literature, topics and methods that underlie a study's theoretical framework. This does not, however, imply that when confronted with an abundance of data, one should ignore well-established theoretical insights. Rather, researchers seeking to introduce innovation into retail geography should be unafraid of generalizing their findings outside existing paradigms, as opposed to leading convoluted efforts at contextualising them amongst the status quo. With these comments in mind, this final section discusses the potential for cross-pollination between machine learning and retail geography for the reinterpretation, reformulation and validation of theories that explore consumer perceptions of retail environments.

Beyond coarse approximations of theory

As shown by Sections 2.1 and 2.2, retail geography is understood with reference to a rich palette of theories. A legacy of research relates the spatial configuration of people to the embedding of attractive retail environments within the urban fabric. Often, however, findings that reveal intuition behind shopping spaces are premised on sample designs that lack adequate coverage, raising issues of external validity and generalizability. As mentioned before, we reiterate that small sample studies remain paramount to scientific enquiry. However, offerings from the changing data landscape and accessibility of machine learning libraries enable refreshing means of testing theories in retail geography. More concretely, unprecedented coverage of characteristics describing urban consumer behaviour enable research beyond the traditionally coarse approximations of phenomena under study. While understandings of consumption spaces are premised on a rich suite of theories, their extent of truthfulness are contingent on the empirical conditions in which these processes are observed. Thus, a data-driven epistemology provides new means of testing or validating theories in retail geography, and observing the extent of (mis)alignment to existing theoretical insight.

Still, when confronted with a myriad of data, there may be some temptation to naively "throw in" all possible variables to the analysis (Dekimpe, 2019). In this instance, letting the data speak for itself *could* reveal some significant insight. From this, the analyst may deduce some theory based on the observed data generating process, possibly without reference to existing knowledge. Indeed, Lamey et al. (2018) argue that retailing theory, in certain areas, may not be sufficiently developed to formulate hypotheses that account for relevant effects, especially in cases involving higher-order interactions between variables. This discussion in big data of which precedes the other, theory or empirics, is particularly salient to the field of retail geography. This is because retail geography is a discipline of complex systems intrinsically difficult to model, owing to the competition, dependencies and collective behaviour of human agents within retail environments. Accounting for these higher-order interactions, however, is increasingly possible with data-driven research able to accommodate the massive size (rows) and high dimensionality (columns) of secondary data sources (Rust, 2019). Dekimpe (2019) argue of the inevitable shift from theory development-focused to data-driven research, but caveat this by claiming "theory will never become obsolete". Worryingly, direction of the latter sees an emergent line of enquiry that is entirely prediction/data driven (Lehmann, 2020). In our context, the use of big data analytics could be used to make predictions for which attributes of retail environments are favoured by consumers visiting the location. Doing so, however, without understanding the processes that drive "why" a prediction is made does little to build understanding of consumer spatial behaviour. So-called black-box prediction models present a risk to the field, as failing to discover the underlying mechanisms driving consumer choices could mean the discipline rapidly becomes a haphazard accumulation of disparate facts (Coveney et al., 2016; Dekimpe, 2019). Echoing Kandt and Batty (2020), as the volume of datasets, generated patterns and big data analytics grows, *theory becomes more, not less important* in the pursuit of transparently identifying plausible causal domains. Ultimately, while a purely data-driven science is suitable for more routine decision-making, introducing these innovations to retail geography seems more appropriate when blended with reference to underlying theories of causes and effects.

2.6 Summary

This chapter has summarised the prerequisite conceptual and empirical knowledge that motivates the underlying research questions proposed within this thesis. In summary, we introduce a data-driven approach of research to explore consumer perceptions of retail environments, but underlie our empirical decisions with existing theory, which is crucial to reconcile (or contrast) new insights with pre-existing knowledge. Rather than a discontinuity with the retail geography literature, we demonstrate how access to new methods, volumes and forms of data should not repel solid theory development. We motivate this position in reference to Coveney et al. (2016), who argues big data require even bigger theories, meaning in each chapter, our model development exercise proceeds by aligning data with theory. Thus, the proposed objective of this research seeks to provide a grounding in which hypotheses in retail geography can be explored through data-driven science. While we leave a comprehensive discussion of each project to the respective chapters, we offer a brief summary below for illustrative purposes.

To begin, Chapter 3 rationalises a data enrichment exercise that creates conditions in which further exploration of research hypotheses relating to consumer spatial behaviour can be explored. Following this, Chapter 4 uses existing research describing place-based attributes of consumption spaces to rationalise hierarchies of retail environments. Finally, Chapters 5 and 6, use street-level imagery of shopping destinations to evaluate theories that suggest visual characteristics of consumption spaces influence consumer experiences, behaviours and preference for particular environments. The following chapters present the outcomes of these exercises in sequential order, with each prefaced by the academic journal the research is published at, or peer reviewed in.

3 — To what extent can machine learning methods enrich data linkage for increasing understandings of retail environments?

N.B. The research presented in this chapter is an adapted version of the publication: Comber, S. and Arribas–Bel, D. (2019) Machine learning innovations in address matching: A practical comparison of word2vec and CRFs. *Transactions in GIS*. 23: 334– 348. https://doi.org/10.1111/tgis.12522

Abstract

Record linkage is a frequent obstacle to unlocking the benefits of integrated (spatial) data sources. Absent of unique identifiers to directly join records, practitioners often rely on text-based approaches for resolving candidate pairs of records to a match. In Geographic Information Science (GISc), spatial record linkage is a form of geocoding that pertains to the resolution of text-based linkage between pairs of addresses into matches and non-matches. These approaches link text-based address sequences, integrating sources of data that would otherwise remain in isolation. While recent innovations in machine learning have been introduced in the wider record linkage literature, there is significant potential to apply machine learning to the address matching sub-field of GISc. As a response, this paper introduces two recent developments in text-based machine learning – conditional random fields and word2vec – that have not been applied to address matching, evaluating their comparative strengths and drawbacks.

3.1 Introduction

Address matching, the process of identifying pairs of records with a spatial footprint, is increasingly required for enriching data quality in wide ranging, real-world applications. With government bodies, businesses and health-care agencies drowning in an ever-increasing deluge of data, a competitive advantage exists in the analysis of integrated data sources as opposed to analysing databases in isolation (Christen, 2012). Yet, in reality, most real-world databases are noisy, inconsistent and replete with missing values. These issues complicate the integration of data. In fact, the acquisition of matched addresses are often key to spatially enabling data used for visualisation or spatial data mining projects (Boulos, 2004). In the address matching context, while geospatial matching is directed by linking the geometric representations of spatial objects (Du et al., 2017), spatial record linkage focuses on resolving text-based linkages between addresses[1].

In absence of unique identifiers that enable direct linking of data in relational database management environments, practitioners have traditionally relied on mathematical linkage techniques broadly divided by deterministic or probabilistic principles (Churches et al., 2002). While deterministic matching consists of generating hand-crafted rule-bases for classification developed from specialist domain knowledge (Oliveira et al., 2016), probabilistic linkage incorporates the varying distributions of a record's attribute values into the assignment of different weights for each field comparison. Weight assignment is related to the frequencies of value occurrences, with stronger weights given to matches for attributes upon which matching is less likely (Blanchette et al., 2013). For address matching, field comparisons might include comparing the street names of an address pair, with more common street names penalized by a lower weighting factor. In this way, resolving text-based

[1] Environment health studies, for example, rely on spatial record linkage to determine whether individuals in residential locations live within defined zones of exposure to hazardous environments (Cayo and Talbot, 2003; Reynolds et al., 2003; Baldovin et al., 2015)

postal addresses to the same address is a *form of geocoding*, where the quality of the match rate is intrinsically tied to the quality of the underlying reference data layer (Goldberg, 2011). Traditionally, address matching has focused on the probabilistic linkage approaches developed by the US Census Bureau in the 1970s (Jaro, 1984).

More recently, record linkage has been permeated by advances in machine learning. In this article, we focus on introducing two particular innovations into the address matching workflow: Conditional Random Fields and word (address) embeddings. Before classification into address matches, input data requires segmentation into feature columns (Churches et al., 2002). The segmentation of postal addresses into attribute columns representing street numbers, street names or zip codes, for example, has been traditionally undertaken using Hidden Markov Models (HMMs). HMMs use statistical induction to predict, from possible arrangements of hypothetical states, the most likely arrangement to have produced the address sequence, and to then label each state by an attribute field (Christen, 2012). For addresses, labels might identify whether the present state represents a street number or street name. A recent innovation for text segmentation tasks has been the use of trained Conditional Random Fields (CRFs) (Lafferty et al., 2001). While HMMs assume the labelling of text sequences are statistically independent of previous outputs, CRFs are conditional by nature, meaning they assume no independence between output labels. Given real-world text sequences such as addresses are represented by interaction and dependencies between words – zip codes are related to city names, for example – it is reasonable to assume CRFs will perform well on a number of real-world text segmentation tasks.

A second innovation relates to the construction of so-called ‘comparison vectors’ that are used for classifying records into matches and non-matches. Comparison vectors are created for each candidate record pair and contain several attributes that describe the text similarity of each pair (Christen, 2012). Traditionally, comparison vectors have been generated using string similarity metrics that measure the text distance between two address fields – the string similarity between two street names such as "Baker Street" and "Bakery Road", for example. Recently, however, advances from the natural language processing community demonstrate methods that map whole words and sentences to vectors in a continuous vector space. *Word2vec* (Mikolov et al., 2013) is one such method that maps semantically and syntactically similar words to nearby points in a vector space, encoding many linguistic patterns and regularities contained within the text. Such methods rely on a theory of language called the distributional hypothesis which states that words appearing in the same context purport similar meaning (Harris, 1954). In the address matching context, one might hypothesise the word vectors generated for two semantically and syntactically similar postal addresses may be correctly resolved to a match.

While recent advances in machine learning have become adopted in the wider record linkage literature (Köpcke and Rahm, 2010; Nasseh and Stausberg, 2016; Ektefa et al., 2011), the address matching sub-field of GISc holds significant potential for the application of machine learning. In this article, we explore how these advances can be integrated into the address matching workflow. In particular, we empirically evaluate the performance of CRFs and word2vec in computing high quality match rates between pairs of postal addresses. The remainder of this article is organised as follows. Section 3.2 introduces the data challenge. Section 3.3 motivates the methodology of the workflow. Section 3.4 presents the findings of the applied address matching methods. Section 3.5 concludes the article.

3.2 Data

Our comparison relies on a set of addresses previously matched by the Local Data Company (LDC) (Singleton, 2015). This provides a ground truth of address pairs with a known match status, which allows us to evaluate the performance of the linkage methods. These address pairs were obtained from a previous round of matching between non-domestic addresses of the LDC and Valuation Office Agency (VOA) databases. In particular, these address pairs reflect matches between LDC records of high street shops to commercial addresses contained in the VOA 2010 rating list (VOA, 2014). A description of the address fields for the LDC and VOA addresses that were segmented by a method we introduce later, the Conditional Random Fields (CRFs), are tabulated in Table 3.1, and are introduced to familiarise the reader with components of a structured address string.

Tag	Description	Example
House	Venue or business name ascribed to the address.	Automotive Solutions
Number	Street-facing building number or apartment number.	43
Unit	A secondary unit designator that identifies an office, unit or apartment.	4a
Level	Expression signifying a floor number.	Ground Floor
Street	Identifying name given to a street.	Paradise Street
Suburb	Unofficial neighbourhood name.	Ropewalks
City	Any human settlement such as the metropolis, city, town or village.	Liverpool
District	Second-level administrative division.	North West
State	First-level administrative division.	England
Zip code	Postal code used for mail sorting.	L3 5TB

Table 3.1: CRFs parser label tags and descriptions identified for the LDC and VOA addresses. *Note*: tags are aligned to address fields of the OpenCage (2018) address formatting library.

Crucially, the matched set of LDC to VOA addresses contains 110,742 pairs that resolve to the same address. This matched set is augmented with 934,150 synthetic non-matched pairs that are generated with the Freely Extensible Biomedical Record Linkage (FEBRL) (Christen and Churches, 2019) Data Set Generator. This works by creating variants of the matched addresses with different error characteristics introduced to the data, meaning the models learn the representations of non-matched addresses (Christen, 2012). Thus, for each address field in Table 3.1, with exception to city and zip code, we introduce error characteristics to the data. A demonstration of how the synthetic non-matches are generated are shown by Table 3.2, where three examples of records from the LDC and VOA database are mutated with different error characteristics. In particular, we set the probability of a missing field as proportional to the number of missing fields in the matched addresses. Moreover, we assign the maximum number of modifications per address field and per address string to one, also testing a scenario where we increase the number of modifications to four later on. These modifications introduce a probability for a character in the address field to be randomly inserted, deleted, substituted or transposed. Importantly, the match status of these synthetic non-matches are always set to false, meaning our machine learning techniques learn the representations of non-matched addresses for highly nuanced cases. To prepare the data for segmentation, we append all address fields from the LDC and VOA datasets into a comma separated address string – 'Home Bargains, 28, Church Way, Bradford', for example – while keeping the zip code separate for reasons we explain immediately below.

ID	House	Number	Unit	Level	Street	Suburb	City	District	State
1-original-LDC	Kwik Fit	67-69	-	-	Whiteladies Road	-	Bristol	South West	
1-duplicate-1	-	41	Unit A	-	White Avenue	-	Bristol	South West	England
1-duplicate-2	Fitness Quick	64-66	-	-	Whiteladies Road	-	Bristol	-	
2-original-VOA	Thomson	-	Unit 19c	-	Teeside Retail Park	Goodwood Square	Stockton-on-Tees	North East	-
2-duplicate-1	-	5	Unit 1a	-	Teeside Retail Park	Goodewood Square	Stockton-on-Tees	North East	
2-duplicate-2	-	-	Unit 5b	-	Teaside Retail Park	Goodwood Square	Stockton-on-Tees	North East	
3-original-VOA	Body Shapers	-	72c	First	Ridgeway Street	-	Plymouth	South West	-
3-duplicate-1	-	-	8a	Second	Ridgeway Street	-	Plymouth	South West	-
3-duplicate-2	Bdoy Shapres	-	72e	-	Longmore Road	-	Plymouth	South West	-

Table 3.2: FEBRL Data Set Generator (Christen and Churches, 2019) demonstration for example records from the LDC and VOA database.

Our next step introduces *blocking* for each method to increase the computational tractability of the linkage task. On a Dell Precision Tower 7000 series with 60GB RAM and multi-core processor, the computational expense for the comparison of each address in the LDC and VOA datasets is substantial. This is because the number of address comparisons without blocking is a function of the Cartesian product of both datasets, which has a quadratic complexity of $O(n^2)$. So, for example, if the LDC and VOA datasets both contained just 10^4 records, the linkage task requires 10^8 comparisons, which becomes computationally non-trivial. To remedy this, we introduce blocking to partition the set of all possible address comparisons between the LDC and VOA databases to within mutually exclusive blocks (Newcombe and Kennedy, 1962). Now, if we let b equal the number of blocks, we are left with n/b addresses per partition on average, which reduces the complexity to $O(n^2/b)$ (Christen, 2012). This means the linkage task becomes tractable even on low-performance machines, as the linkage within each partition can be processed sequentially or, alternatively, in parallel if the user has access to a multi-core machine. Therefore, in each of the following methods, we use the zip codes of postal addresses as a blocking key. This reduces the number of address pair comparisons to within 39,855 zip code 'blocks', with the distributional characteristics of these partitions displayed in Figure 3.1. In our case, the near uniform frequency distribution of the zip code blocks and completeness of the zip code attribute means it is a sensible choice as a blocking key. Yet, for different address databases where the zip code column is replete with missing values, an alternative attribute should be considered as a blocking key, which is an empirical decision to be motivated by the characteristics of the databases's attribute columns.

Yet, one potential issue with using zip codes as a blocking key are typographic errors in the spelling. In our case, this is pertinent because while validation checks are employed by the LDC, the recording of commercial addresses are undertaken by teams of surveyors, and are therefore susceptible to human error. To account for this, we explore *sorted neighbourhood blocking*. We sort together the LDC and VOA datasets using the zip code

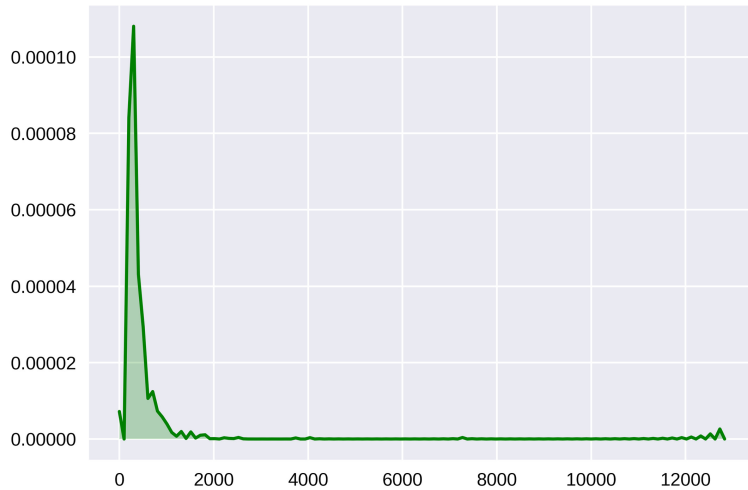


Figure 3.1: Distributional characteristics of the address block partitions ($n = 39,855$).

as a sorting key value while restricting address comparisons to only records within a window of fixed dimension, $w = 5$. As the window slides over the sorted zip codes, the identification of matches and non-matches are restricted to candidate address pairs within this window of fixed size (Cibella and Tuoto, 2012). This means the technique is highly sensitive to lexicographic order which, in our case, is advantageous because we create candidate address pairs from misspelt zip codes – for example, the comparison between addresses with sorted zip codes 'L3 5TA' and 'L3 5TB' that pertain to the same address. However, a limitation of sorted neighbourhood blocking is that misspellings of the first character in the blocking key can erroneously direct addresses to incorrect zip code blocks. Q-gram blocking is an alternative approach to generate partitions of zip code blocks, and converts blocking keys to a list of bigrams ($Q = 2$). The zip code "SW11 9LU", for example, generates a bigram list as ['SW','W1','11','19','9L','LU'], from which k sub-lists of length $k - 1$ are generated recursively to create variations of the zip code (Christen, 2012). This continues up to a minimum threshold for the number of bigrams in the sub-list. Following the example, this might create blocking key values such as 'W1 9LU', 'SW1 9LU' and 'SW19 LU', meaning that the same address is likely to be inserted into several different zip code blocks. While the recursive generation of sub-lists is computationally expensive, the advantage of Q-gram blocking is that it overcomes typographic error in the characters of the blocking key. In our case, given the zip code attribute was profiled as generally being of high quality, the choice of blocking mechanism was less fundamental. Nevertheless, we introduce additional blocking techniques to provide instructional guidance to researchers interested in replicating our methodology on their own datasets, where attributes of the blocking key values might be less clean.

3.3 Methods

3.3.1 Conditional Random Fields

One of the principle challenges in obtaining high quality match rates is the conversion of raw data into a structured, usable format for comparison. For postal addresses, this involves parsing address sequences into feature columns. Take, for example, a canonical address of the form '3B Records, 5 Slater Street, Liverpool L1 4BW'. Our objective is to segment (or label) this address into appropriate columns for business name, property number, street name, city and zip code. To use a Hidden Markov Model (HMM) (Baum and Petrie, 1966) for segmentation would be to assign a joint probability to the observation sequence where the labelling of address elements are independent of previous labels (Churches et al., 2002). This means, following the example above, '3B' could be incorrectly classified as a property number, whereas it actually completes the business name '3B Records'. Importantly, '3B' will now be considered a property number, and that (alternative) fact will be used to classify the next token, 'Records'. This leads to an erroneous sequence of label predictions.

In real-world text sequences such as addresses, the probability of a transition between labels might depend not just on the current address element, but also on past and future elements. For this reason, Conditional Random Fields (CRFs) (Lafferty et al., 2001) are more suited to address segmentation tasks. Principally, this is because CRFs negate what is known as the *label bias problem*: "transitions leaving a given state to compete only against each other, rather than against all transitions in the model (Lafferty et al., 2001)." Returning to the example, when the CRFs has parsed '3B' and reaches the second token, 'Records', the model scores an $l \times l$ matrix where l is the maximum number of labels that can be assigned by the model. In L, element l_{ij} reflects the score for the probability of the current word being labelled i, and the previous word labelled j (Diesner and Carley, 2008). Returning to the example, when the parser gets to the *actual* property number, '5', the highest score in the matrix indicates the current label should be revised to a property number, and the previous label to a business name. Below, for example, we provide an illustrative example for an erroneous sequence of labels predictions that hypothetically may have been segmented with a HMM.

3B	*Records*	*Slater Street*	*Liverpool*	L1 4BW
NUMBER	STREET	SUBURB	CITY	ZIPCODE

While in the CRFs, prediction of the most likely sequence of labels uses a reversible highest *scoring* path. This is known as Viterbi inference (Viterbi, 1967), and leads to a sequence of labels with the highest likelihood.

3B Records	*5*	*Slater Street*	*Liverpool*	L1 4BW
HOUSE	NUMBER	STREET	CITY	ZIPCODE

With the CRFs model, the previously raw and unstructured address will now be correctly segmented into the following feature columns that can be used as a basis for classifying records into matches and non-matches:

House:	3B Records
Number:	5
Street:	Slater Street
City:	Liverpool
Zip code:	L1 4BW

In our case, address segmentation is undertaken using the *Libpostal* C library (Barrentine, 2013) that trains a CRFs model on addresses sourced from OpenStreetMap (OSM) data. This means we apply a pre-trained address segmentation model to each raw address string, setting the country code of the parser to "GB" (for Great Britain) so the software recognises it is segmenting UK addresses. Libpostal's *parse_address* command will then label the address sequence into features columns, if they exist, for the fields in Table 3.1. To empirically evaluate the performance gain, we also introduce a HMM[2] alongside the CRFs parser. Once feature columns consisting of address elements have been obtained for every LDC and VOA address, a *comparison vector* is constructed for each candidate address pair. Comparison vectors contain several attributes that describe the text similarity between each feature column (Christen, 2012). In our case, each element of this comparison vector contains the Jaro-Winkler string similarity between each address field, with exception to zip code, in Table 3.1. Briefly, Jaro-Winkler distance calculates the minimum number of single character transpositions required for converting one string into another, also increasing the similarity when the first few characters are the same (Herzog et al., 2007). We motivate the decision to use Jaro-Winkler as our string comparison function because previous findings show it performs best on attributes containing named values – property names, street names, or city names, for example (Christen, 2012; Yancey, 2005). If a given address field is missing, the Jaro-Winkler similarity between the pair of address

[2]Our HMM is trained on the same OSM addresses for the UK as Libpostal. They are obtained by filtering the 'great-britain-latest.osm.pbf' file available from Geofabrik (2020). Filtering is performed using the *Osmosis* command line application (OpenStreetMap, 2018) that allows us to distil addresses from the entirety of OSM data in the file. Therefore, we use Osmosis to filter by the following tags: 'addr:housename', 'addr:housenumber', 'addr:street' and 'addr:postcode'. Implementation for the HMM model is provided by a script available from FEBRL (Christen and Churches, 2019), which tags free-text address sequences from look-up tables before rearranging the tags to the most likely sequence of address fields.

fields is set to zero. These comparison vectors for each address pair are the basis of a binary classification for classifying whether address pairs resolve to matches or non-matches. The general idea is that the more similar two addresses are, as described by Jaro-Winkler similarity, the higher likelihood they resolve to the same address.

Our classification approach is supervised, meaning we use our training data of known true match and true non-match status generated in Section 3.2 to evaluate the outcome of our address matching exercise. By training a classifier, we allow the model to learn the nuances of matched and non-matched addresses. This means after training, we can test whether unseen address pairs for which the match status is known correctly resolve to matches and non-matches, allowing us to evaluate the performance of our linkage techniques. Thus, once comparison vectors have been generated for each training record, we introduce several classifiers to facilitate the linkage into matches and non-matches. In particular, two ensemble methods for classification, a random forest (Breiman, 2001) and gradient boosted classifier known as XGBoost (Chen and Guestrin, 2016), are trained alongside a logistic regression model. We motivate our use of ensemble classifiers for one key reason. Our comparison vectors are embedded to a 9-dimensional vector space. This means each dimension reflects one of the nine Jaro-Winkler similarities between each address field in Table 3.1, with the exception of zip codes which are used for blocking. When partitioning matches from non-matches in this 9-dimensional vector space, while the logistic model searches for a linear decision boundary, the multiple decision trees of the ensemble methods partition the vector space into half spaces by using axis-aligned linear decision boundaries (Efron and Hastie, 2016). This has the net effect of a non-linear decision boundary, which is desirable if the comparison vectors cannot be accurately separated into matches and non-matches by a single hyperplane. From here, the classifiers are trained using k-fold cross-validation where $k = 10$, which we explain in Section 3.4, and are evaluated against metrics commonplace in machine learning such as precision and recall (Christen, 2012).

3.3.2 Word embeddings

In our second approach, we augment the use of CRFs with so-called 'word embeddings', which are the name given to the vector representations of words. Vector space models *embed* words in a continuous vector space, where words with similar syntactic and semantic meaning are mapped, or embedded, to nearby points (Mikolov et al., 2013). Such methods leverage the distributional hypothesis of language which states that "words which are similar in meaning occur in similar contexts" (Rubenstein and Goodenough, 1965). One such method is word2vec, a neural probabilistic language model whose training objective is to find word vector representations that are good at predicting the surrounding words in a text-based sentence or document[3]. In our case, we hypothesise that learning high-dimensional vectors from postal addresses may be used to match addresses that resolve to the same geographic location despite irregularities in the text. In practice, we train *gensim's* (Řehůřek and Sojka, 2010) implementation of word2vec on 29.6 million parsed postal addresses from the UK Postcode Address File (PAF) database (PAF, 2018). Learning word embeddings using word2vec requires setting the dimensionality of the vectors, so the training phase begins by randomly initializing each address component with 100 real numbers. Following the literature (Mikolov et al., 2013), we set the dimensions of the word vectors generated for the parsed address fields to 100, meaning each field is represented by an array of numbers of length 100 which, as an example, can be represented as: $\begin{bmatrix} 0.32 & 0.28 & ... & 0.01 & 0.58 \end{bmatrix} \in \mathcal{R}^{100}$. By feeding successive address fields to the model, the real numbers of each word vector are updated so that words sharing the same context are mapped closer together in the vector space. To build intuition towards this idea, we employ Figure 3.2, where a t-SNE (van der Maaten and Hinton, 2008) dimensionality reduction technique is applied to the top ten closest vectors to an address field for property name, "halifax plc". In this 2-dimensional vector space, the closeness of word vectors, measured

[3] For details of technical implementation, the reader is referred to Mikolov et al. (2013).

by cosine similarity, represent words that share closer semantic and syntactic meaning. 'Halifax PLC', for example, is a bank, and interestingly the word vector generated for it is embedded nearby to business's that have a financial remit, 'Natwest', 'TSB Bank' and 'Barclays Bank PLC', for example.

Figure 3.2: t-SNE visualisation demonstrating the top 10 closest vectors to the word "Halifax PLC" in a 2-D vector space.

Under this approach, instead of using Jaro-Winkler similarity between the address fields, we augment the linkage task by comparing word vectors generated by word2vec for the address fields segmented by the CRFs model. Thus, after training a word2vec model on PAF addresses, for every LDC and VOA address we are able to obtain a 100-dimensional vector for each address field. The postal address, '5, Myrtle Street, Liverpool', for example, contains three address fields – a street number, a street name, and a city name – we obtain vectors for. Similar to our first approach, we construct a comparison vector, but this time each element is the *cosine similarity* between the word vectors constructed from the address fields parsed by the CRFs model. In cases where address fields are missing, we set the cosine similarity to 0, which infers orthogonality or linear independence between the vectors under comparison. The decision to choose cosine similarity as the criterion for measuring similarity between address fields is because it has favourable qualities in capturing the semantic closeness of word vectors (McInnes and Pedersen, 2013). As before, we train a random forest, XGBoost and logistic regression model on the comparison vectors and associated match status labels using k-fold cross-validation to evaluate the linkage performance.

3.4 Results

To evaluate the performance of the HMM and CRFs alongside the CRFs augmentation with word2vec, we use address pairs for which the match status is known (as discussed in Section 3.2). The results of these methods are highlighted in Table 3.3, and are benchmarked by evaluation metrics known as *recall* and *precision*. Recall measures the proportion of address pairs that should have been classified, or recalled, as matched (Christen, 2012). The precision – or, equivalently, the positive predictive value – calculates the proportion of the matched address pairs that are classified correctly as true matches (Christen, 2012). To minimize over-fitting our supervised models, we introduce k-fold cross-validation where $k = 10$, meaning the training data is split into ten disjoint groups. In each split, the classifier is trained and tested on these subsets of address pairs, with the resulting recall and precision averaged across the groups. This means for each group we have a randomized training and testing set split by 75% and 25%, respectively.

To begin interpretation, we first turn attention to the baseline HMM that we use as a point of comparison for the machine learning techniques we introduce. Consistent with our earlier motivations, when address fields are parsed with the CRFs model, they outperform the HMM. This is shown by the lower recall values retrieved by each of the classifiers using

Table 3.3: Recall and precision evaluation metrics for the HMM, CRFs, and CRFs augmented using word2vec.

Method	Precision	Recall
HMM		
Logistic	0.738	0.459
Random forest	0.944	0.696
XGBoost	0.959	0.688
CRFs		
Logistic	0.933	0.820
Random forest	0.940	0.918
XGBoost	0.955	0.902
CRFs-Word2vec		
Logistic	0.870	0.687
Random forest	0.933	0.874
XGBoost	0.950	0.870

† *Note*: results are 10-fold cross-validated using 25% of the data for testing within each fold.

the HMM technique. Interestingly, the precision values of the HMM and CRFs techniques are broadly consistent. This implies that, of the total number of matches returned, both techniques perform well at partitioning true positives from false positives, but the CRFs classifies a larger proportion of matches as shown by the higher recall value. This finding suggests, unlike the HMM, that the reversible sequence of labelling introduced by the CRFs leads to higher quality match rates. We now turn direct attention to the supervised classifiers that are trained on the comparison vectors built using the CRFs model. An interesting facet of tree-based models is that the feature importances can be recovered (Hastie et al., 2009). The 'importance' of different features, or, equivalently, address fields, to the match classification are visualised by the red bars in Figure 3.3, along with the inter-trees variability. Figure 3.3 is consistent with conventional wisdom, as it indicates the street name and house number are the most important features that are used when resolving candidate pairs of addresses to a match. To visualise the absolute numbers of matches, we provide a confusion matrix in Figure 3.4 for the random forest trained with address pairs parsed by the CRFs. In Figure 3.4, the top-left quadrant shows true negatives, top-right shows false positives, bottom-left shows false negatives, and bottom-right shows true positives. Briefly, true positives are address pairs labelled as matches that are true matches; false positives are address pairs mislabelled as matched; true negatives are records classified as non-matched which are true non-matches; and false negatives are addresses classified as non-matched but are actually true matches (Christen, 2012).

From Table 3.3, it is clear the ensemble learners offer slight improvement over the logistic model in returning a larger fraction of true positives amongst all returned 'matches'. This is shown by the marginally higher precision value for the random forest, which we complement by displaying a precision-recall curve in Figure 3.5. The green line in the figure suggests that while the random forest performs well at classifying true matches from the matches it returns, it performs less well at retrieving all matched instances. Between the two ensemble approaches, XGBoost classifies the highest number of true

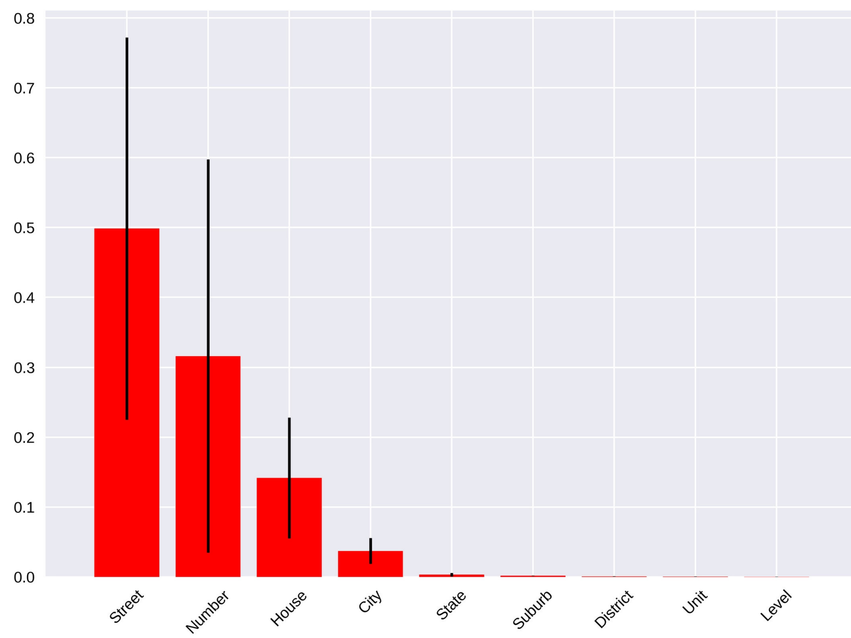


Figure 3.3: Feature importances of address fields from Table 3.1 to matching outcomes. Importances are given for the random forest model trained on address fields segmented by the CRFs.

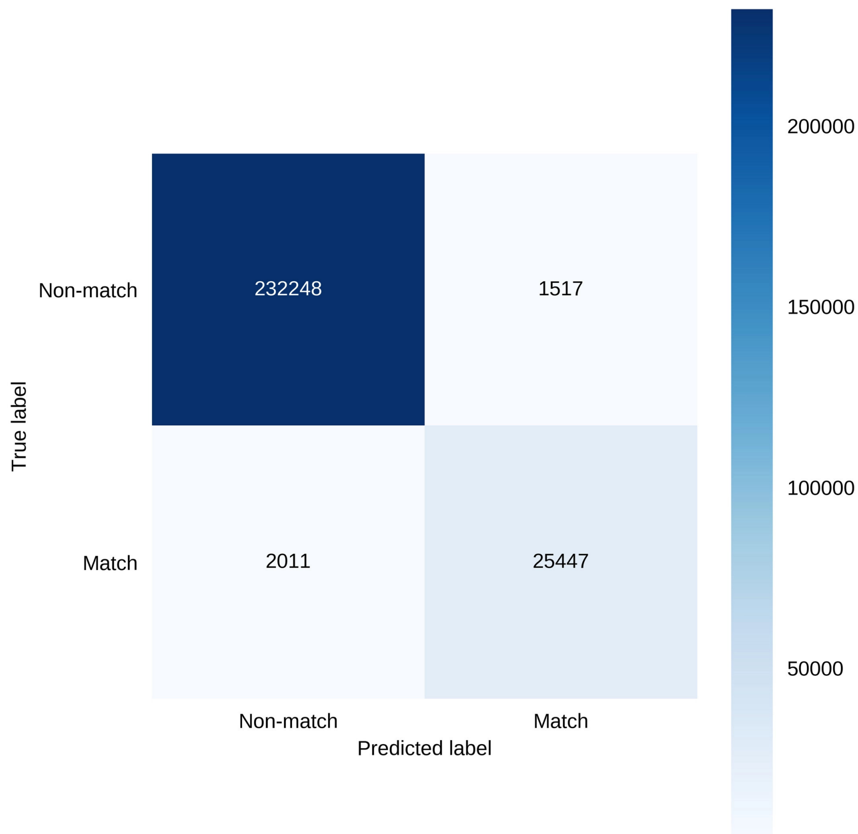


Figure 3.4: Confusion matrix for true positives, false positives, true negatives and false negatives retrieved by the random forest classifier for CRFs segmentation.

matches correctly, with there only marginal differences between the recall, or retrieval of relevant address pairs, between the two. Presumably, the ensembles perform better because the vector space for classifying address pairs is not linearly separable, and requires a non-linear decision boundary to partition matches from non-matches to a high degree of accuracy.

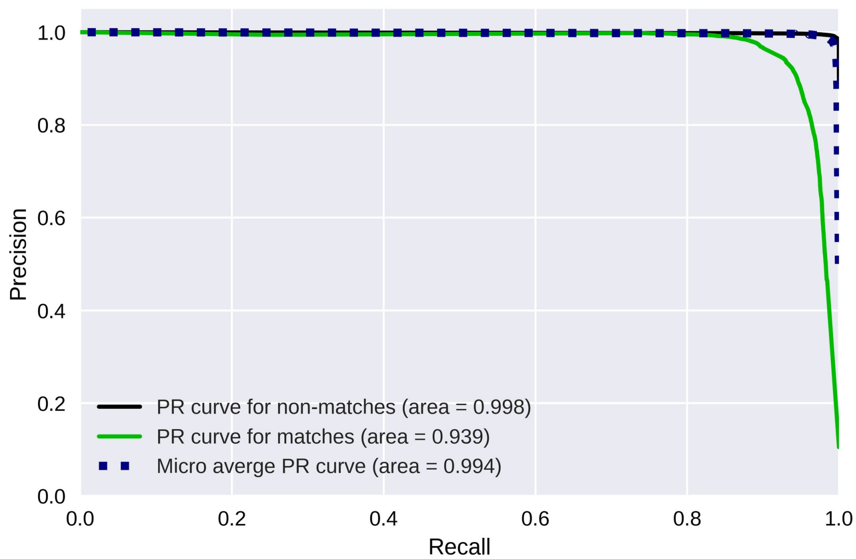


Figure 3.5: Precision-recall (PR) curve for the random forest classifier trained on addresses segmented by the CRFs model.

Next, we turn attention to the CRFs method that is augmented with the use of word2vec for address field comparisons. While the first method uses Jaro-Winkler similarity to assess the string distance between address fields parsed by the CRFs model, our second approach augmented the first by replacing Jaro-Winkler similarity with cosine similarity between word vectors learnt from the parsed address fields. Overall, this augmented approach is highly competitive with the first approach. If, for example, we take XGBoost findings from Table 3.3 as a point of comparison, the precision and recall values only decrease by 0.005 and 0.032, respectively, in the augmented approach when compared to the first approach. In all, the findings support our theoretical motivations from Section 3.3, and imply the word vectors obtained for the address fields perform well in capturing the syntactic structures and regularities required to resolve address pairs to a match.

Finally, we tweak two components of our empirical design to evaluate the robustness of our main findings. In doing so, we evaluate any substantive change in the match performance for our preferred specification, the XGBoost classifier, trained on comparison vectors generated by the CRFs model. Firstly, we adjust the probabilities for introducing modifications to the synthetic non-matched addresses generated by FEBRL. Here, we apply two scenarios: in one, we set the probability of a missing field equal to the proportion of missing fields in the matched addresses, while setting the maximum modification per field and maximum modification per address to one; and in two, we maintain the same probability for missing fields, but increase the maximum number of modifications per field and maximum modifications per address to nine in both cases. Modifications refer to insertions, deletions and transpositions of characters in the address field. These scenarios are introduced to evaluate the extent to which the classifier performance changes as we degrade the quality of our non-matches in the training data. As expected, the precision value increases marginally from 0.955 to 0.973 as we increase the number of error modifications to the non-matched addresses. This is intuitive as increasing error in the non-matches means the address pairs become more dissimilar. Therefore, it becomes easier for the classifier to disambiguate between matches and non-matches as the nuances between non-matched address pairs become less pronounced. Our second tweak tested the change of blocking mechanism from standard blocking to sorted neighbourhood blocking and Q-gram blocking. This was applied to better handle cases of misspelt zip codes, which is problematic because misspellings could allocate addresses to incorrect zip code blocks. Despite the empirical motivations to alternate the blocking mechanism, we found the results of our main findings were invariant to which blocking technique was applied. Presumably this was because zip codes of the LDC and VOA addresses are of high quality, as the VOA addresses originate from an official UK government source and the LDC have a business case in maintaining accurate, high quality zip codes.

3.5 Conclusion

Often the biggest problem when faced with spatial data is accessing it. Address matching resolves text-based address sequences to matches, integrating disparate sources of data that would otherwise remain in isolation. In this article, we evaluated the performance of two recent machine learning techniques for linking address pairs where the match status was already known. Our first technique, the CRFs approach, focused on segmenting whole postal addresses into address fields, which became the basis for constructing comparison vectors for every candidate address pair in the dataset. Once obtained, supervised classifiers were applied to partition the comparison vectors of address pairs into matches and non-matches. In all, the classifiers trained using addresses segmented by the CRFs achieved a precision of up to 0.955, with the ensemble learners outperforming the logistic model. This was likely due to the improved fit of a non-linear decision boundary to the underlying vector space.

Our second approach augmented the first, by replacing the string similarity metric we used to compare parsed address fields from the CRFs model with a comparison between word vectors. These were generated using a technique called word2vec, which sought to embed semantically and syntactically similar address fields to nearby locations in the vector space, with the expectation that vectors embedded *nearer* together could be used to match address fields. As before, we used supervised classifiers to facilitate the linkage, which resulted in a precision of up to 0.950. This value implied the vectors obtained for the address fields performed successfully at encoding word relationships, patterns and regularities that are required to facilitate the accurate linkage between address pairs. In synthesis, the main implications of this article point to the utility of CRFs, and its augmentation with word2vec, for the accurate segmentation of addresses. These steps are preconditions for constructing high-quality comparison vectors that can be used to accurately classify address pairs into matches and non-matches.

4 — Do urban hierarchies reflect spatial configurations of attractive consumption spaces and retail agglomerations?

N.B. The research presented in this chapter is an adapted version of the publication: Comber, S. Arribas–Bel, D., Singleton, A., Dong, G. and Dolega, L. (2019) Building Hierarchies of Retail Centers Using Bayesian Multilevel Models, *Annals of the American Association of Geographers*. https://doi.org/10.1080/24694452.2019.1667219.

Abstract

The perceived quality of urban environments are intrinsically tied to the availability of desirable leisure and retail opportunities. In this paper, we explore methodological approaches for deriving indicators that estimate the willingness-to-pay for retail and leisure services offered by retail centres. Most often, because the quality of urban environments cannot be qualified by a natural unit, the willingness-to-pay for an urban environment is explored through the lens of the residential housing market. Traditional approaches control for individual characteristics of houses, meaning the remaining variation in the price can be unpacked and related to the availability of local amenities, or equivalently, the willingness-to-pay. In this paper, we use similar motivations but exchange housing prices for residential properties with property taxes paid by non-domestic properties to glean hierarchies of retail centres. We outline the applied methodological steps that includes very recent, non-trivial contributions from the literature to estimate these hierarchies, and provide clear instructions for reproducing the methodology. Using the case study of England and Wales, we undertake a series of econometric experiments to rigorously assess retail centre willingness-to-pay (RWTP) as a test of the methods reviewed. We build intuition towards our preferred specification, a Bayesian multilevel model, that accounts for the possibility of a spatial autoregressive process. Overall, the applied methodology describes a blueprint for building hierarchies of retail spaces and addresses the limited availability of spatial data that measure the economic and social value of retail centres.

4.1 Introduction

The quality of an urban environment is a principle determinant of attractiveness (Glaeser et al., 2001). Attractiveness, in this context, might be understood as an outcome of perceived place attributes (Finn and Louviere, 1996), which can be argued as those perceptions, attitudes and patronage behaviour of consumers drawn to particular places (Teller and Elms, 2012). However, the quality of an urban environment cannot be qualified by a natural unit of analysis, and so approaches typically observe its capitalisation into housing prices (Rappaport, 2009). The depth and breadth of consumer amenities, natural and cultural assets, and opportunities in the labour market are seen as an influential driver of demand for residential space (Oner, 2017). As an example, the attractiveness of Paris might be considered as a product of fine-dining restaurants, art museums such as the Louvre, and the impressive stock of buildings (Brueckner et al., 1999). Accordingly, Rappaport (2009) argues that environments with above-average consumer amenities, or implicitly, quality of urban environment, typically sustain a higher density of residential population, resulting in higher prices in the housing market.

Under these assumptions, the desirability of areas has often been explored through the lens of home-buyer decisions in the residential housing market. Hedonic analyses that estimate the willingness-to-pay for consumer amenities through residential housing markets derive a snapshot for the desirability of particular places. In recent years, the proportion of the individual's spend allocated to consuming the economy's lifestyle amenities and services has increased substantially (Oner, 2017). An increasing share of the individual's rising wealth is allocated to the pursuit of enjoyment and experience, which is reflected by an increase in the willingness-to-pay for properties that are proximate to retail and leisure destinations. Changing consumer desires have transformed traditional retail zones into spaces of leisure consumption that are increasingly service-orientated. Concentrations of retail outlets are referred to as retail agglomerations and exist across a system in space, with their attractiveness to home-buyers related to the composition and richness of the retail environment, but also competing opportunities available elsewhere (Teller and Elms, 2012). Moreover, areas of retail perform as attractors for a multitude of heterogeneous user groups such as prospective and existing residents, consumers, visitors and employees (Teller and Reutterer, 2008). In this way, the availability of consumer amenities are seen as a driver of urban vitality, and so an estimation of the willingness-to-pay for an amenity-rich environment can be used to gauge how desirable that area is.

One particular area that attracts a number of retail opportunities is the town centre. Town centres are complex urban economic systems that are characterised by the clustering of socio-economic activity (Thurstain-Goodwin and Unwin, 2000). Embedded within the urban fabric of town centres are retail centres which are agglomerations of consumer spaces and shopping destinations that are central to economic and civic life (Pavlis et al., 2018). Town centres are typically comprised of a retail centre, but in some cases have more expansive functional areas that include office spaces in addition to retail and services. A focus on classifying *retail centre* willingness-to-pay (RWTP) is foundational to understanding hierarchies of retail spaces which, by implication, reveal geographic patterns in urban growth and development. Retail centre hierarchies are the rankings of particular centres within a network, whose position relates to the size, attractiveness and gravity of their composite retailers influence, with top ranked centres typically offering multi-purpose, comparison shopping experiences that have a wider geographical reach on consumers (Dennis et al., 2002). By contrast, smaller district centres are more embedded in local economies and are patronised by a smaller catchment area. While an underlying driver to the sustainability of the built environment, since the 1970s retail centres have become threatened by the decentralisation and dispersal of development to out-of-town locations on the periphery of towns. Not only this, Singleton et al. (2016) claim retail has become increasingly vulnerable to the effect of growing online shopping, and so must be considered within a framework of e-resilience.

In the present paper, we introduce a statistical technique to derive indicators that describe hierarchies of retail centres across the national extent, which we obtain alongside a measure of uncertainty in the rank-ordered estimate for each retail centre. Despite the concerns raised above, while retail centres in the UK have long been examined under a series milestone reviews (Department of the Environment Urban and Economic Development Group, 1994), there is little quantitative evidence that explore the performance of town centre retail economies, which has undermined effective policy formulation and decision-making (Astbury and Thurstain-Goodwin, 2014). Indicators of retail hierarchies produced by commercial organisations (CACI, 2018), for example, lack fine spatial granularity at the retail centre scale. Our approach is motivated by a hedonic framework of analysis that is typically orientated towards residential housing markets, except that we exchange residential for commercial properties to execute our empirical strategy. We describe the methodological steps required to reproduce the RWTP estimates, which includes very recent, non-trivial contributions from the econometrics literature. Finally, we introduce a validation exercise to verify the RWTP estimates correspond to conventional wisdom by correlating the scores to socio-economic characteristics of the retail centre. Not only is the approach we operationalize novel in application, but we note our methodology is replicable and generalisable to international contexts, conditional on data availability.

The remainder of the paper is organised as follows. The next section motivates the underlying conceptual framework of the paper, followed by an introduction to the specification and underlying assumptions of the modelling approach. After elaborating on the nature and limitations of the data source, we step through the results of each model, including a validation exercise to confirm whether the RWTP estimate for each retail centre responds to characteristics that are associated with attractive places. The final section summarises the article, presenting extensions for future elaborations of the applied methodology.

4.2 Background and Motivation

4.2.1 Modern Consumption Patterns

The desirability of urban places to live is increasingly dependent on their ability to provide consumption opportunities, which are often reflected in housing prices (Glaeser et al., 2001). Leisure and retail amenities such as restaurants, live performance venues and shopping districts, have been shown to be crucial for attracting modern day workers who balance economic *and* lifestyle opportunity in selecting places to live and work (Florida, 2000). As perceptual qualifications for the quality of leisure and retail environments cannot be directly counted nor observed, it has often been evaluated by the willingness-to-pay for residential property through hedonic approaches (Rivera-Batiz, 1988; Hui and Liang, 2016). Jin and Sternquist (2004) argue the desire for leisure and shopping is increasingly linked to the concept of enjoyment and experience. From a consumer perspective, shopping trips not only satisfy the individual's bundle of wants and needs at a given store, but they allow the consumer to speak their own geographies of everyday life through the language of consumption (Sack, 1988). This "credit-card citizenship" towards identities and preferred lifestyle choices provides opportunity for social mixing and participatory entertainment (Goss, 1993). Over the last few years, however, this traditional brick-and-mortar retailer landscape has been restructured by the growth of electronic retailing, with e-commerce sales in the U.S. rising by 101% in the period between 2011 to 2016 (Helm et al., 2020). Due to the rise of the internet, online consumption has tilted power from retailers to consumers through opportunities for 24/7 convenience and price comparison, increased ease of market entry and transparency, and also a distribution of products to a wider geographical reach (Williams, 2009). Evidence suggests this rapid expansion in online consumption has impacted the health of retail centres in complex ways, and has been a principal driver of change to the geography of traditional UK high streets (Wrigley and

Lambiri, 2014).

Adjustments as a result of online shopping to the market share of retailing, leisure and services in retail centres are typically considered by detrimental effects that cause physical shopping opportunity to be substituted online (Doherty and Ellis-Chadwick, 2010). Yet, online retailing has also been linked to complementarity and modification processes that blend traditional retail channels with e-commerce by refashioning the in-store consumer experience (Poushneh and Vasquez-Parraga, 2017). In the UK, major retailers including Argos, John Lewis, and Boots have integrated new technologies by opening 'click and collect' points that act as points of delivery for internet sales by allowing customers to order goods online and collect them in-store (Singleton et al., 2016). Thus the role of retail centres remains vital to modern consumption and the continuity of physical shopping environments, with consumers pointing to the hedonic experience physical stores offer through recounted social experiences, the opportunity to discover new and exciting goods, and the gratification afforded by touching or trying products in-store (Cho and Workman, 2011). Under this lens, Singleton et al. (2016) recast the propensity of localised populations to engage with the mixture of online shopping and physical retailing provision under a frame work of 'e-resilience'. The constraint *or* opportunity of e-commerce to retail centres are not uniform across all retail types, with retailers whose merchandise can be replicated and digitised online the most vulnerable to large-scale store closures and lost physical shopping opportunity (Zentner et al., 2013).

4.2.2 Geographic Behavioural Drivers of Retail Centre Hierarchies

More concretely, the geodemographic characteristics of catchments served by retail centres are fundamental drivers of consumer choices and behaviours that shape the willingness-to-pay for retail opportunity, and, in turn, hierarchies of retail spaces (Birkin et al., 2002). In the UK, geographic variation of consumer disposable incomes affect the relative retail value of catchment areas. For example, hierarchies of retail centres for large conurbations and metropolitan centres are moderated by their propensity to attract highly mobile consumers who require multiple retail and leisure choices (Wrigley et al., 2015). More generally, steps in the hierarchy of retail centres have become contingent on a rising 'convenience culture'. This incorporates the progressive rise of online retail with preferences for 'local' shopping (and derived product authenticity, traceability and sustainability benefits) alongside a revaluation of consumer awareness towards 'community-sustaining' consumption (Chalmers et al., 2012). Since the early 2000s, significant demographic and societal shifts have driven these trends, with particular growth among low density households, ageing populations, and younger workers who are faced with longer working hours and busy lifestyles (Wrigley et al., 2015). These groups in particular have an increasing desire for convenience at the local level. In the UK this is revealed by evidence from the Institute of Grocery Distribution (IGD) that suggests consumers are increasingly shopping little and often at shops closer to home rather than shopping at larger out-of-town retail developments, a phenomena described as 'top-up shopping' (IGD, 2014). Moreover, a report by the Ethical Consumers Market suggests the number of shoppers purchasing produce from local shops increased from 15% to 42% between 2005 and 2012 (Ethical Consumer Research Association, 2013). This has considerable beneficial implications for the configuration of the UK's retail centres, as high streets and town centres are now increasingly the preferred locations for consumers to undertake their *top-up* shopping. Not only has this driven footfall back to retail centres, but local shopping has reshaped hierarchies of retail spaces by boosting the vitality and viability of town centres and high streets in the UK (Wrigley et al., 2015).

Yet there is significant demographic variation in the propensity for consumers to value local shopping and engage with internet retail; this has determined the differential geographies of online shopping (Longley and Singleton, 2009) and, in turn, been an influential driver of retail hierarchies. Whilst typically younger age groups have been the most receptive to online shopping, significant growth has been recorded in the rate of online purchasing amongst those aged 65+, with 48% buying online in 2014, increasing from 16% in 2008 (ONS, 2018). By exploiting opportunities provided by digital technologies and adapting retail spaces to the meet the needs of every population group retail centres have become virtual marketplaces. Here, consumers are able to access information online regarding the availability of products, stores, services and brands prior visiting which has enhanced the retail centre customer experience (Wrigley et al., 2015). However, despite these significant structural changes, good product ranges, quality of retail provision and traditional factors such as overall retail centre experience, atmosphere and leisure provisions remain foundational drivers of footfall in retail centres. This extends their use from shopping destinations to areas for economic and educational activities (in addition to social interaction) (Warnaby et al., 2002).

In addition to demographic variation, consumer behavioural patterns also vary spatially and are directly linked to the geographies of demand towards retail facilities. Steps in the hierarchy of retail centres are intertwined with the underlying characteristics of the catchment area itself. Variations in consumer confidence, the ownership of basic digital skills and local supply factors such as convenience and accessibility at the small area level are influential factors towards the vitality of retail centres (Wrigley and Dolega, 2011). Thus, the propensity and desirability of consumers to engage with physical shopping opportunity is governed by a multitude of contexts and influences such as: the rurality and remoteness of an area (Warren, 2007), the extent of internet connectivity and speed of connection (Singleton et al., 2016), and even how informed (and educated) consumers are to access online retail (Helsper and Eynon, 2010). Despite these factors, and even in a digitally transformed retail landscape, the demand for high street shops remains a permanent fixture of consumer desires, and so an estimation of the willingness-to-pay for retail centres is foundational to unpacking hierarchies of retail spaces that reveal geographic patterns in urban growth and development.

4.2.3 Measuring Attractiveness

Within the academic literature, measures for estimating attractiveness[1] are most typically classified into two streams of research. Models of the *first* stream are inspired by Reilly (1931)'s gravitational law of retail, which motivated the seminal work of Huff (1963). The Huff Model applies Newtonian laws of physics to estimate a retail catchment area that factors the spatial distribution of competing retail destinations when evaluating their gravity or consumer pull to different population groups (Dolega et al., 2016). Huff models are advantageous because they simultaneously estimate break points in the demand surface for all competing retail destinations in the model, and reduce the probability of a consumer to patronize a given location to three groups of variables, namely distance between shops and consumer's homes, a measure of attractiveness such as store size, service levels, or opening hours, and competition proxied by the number of retail units in a location (Teller and Reutterer, 2008). Yet, the usual criteria for retail attraction in Huff models are often argued as incomplete, as additional factors that impact the consumer's propensity to visit a retail destination involve a suite of qualitative indicators. These include: the variety of retail-tenants; site-related factors such as accessibility and parking conditions; and environmental factors reflected by sensual stimuli such as ambience, atmosphere and perception of safety (Teller and Elms, 2010). Clearly, these indicators influence the choice of shopping destination, but measuring this is difficult to obtain across a national extent (Dolega et al., 2016).

Methods of the *second* stream are motivated by findings that demonstrate housing

[1] Given this paper's intersection between retail geography and urban economics, particular attention to the conceptualisation of *attractiveness* is required. While urban economists perceive attractiveness through an estimation of willingness-to-pay, retail geographers might observe the attractiveness of shopping environments through a lens of image-based characteristics such as cleanliness of the shopping environment, plurality and variety of shops, or existence of fun and entertainment programs (El–Adly, 2007; Chebat et al., 2010; Gomes and Paula, 2017). Thus, to avoid confusion, in the present paper we adopt the direction of the former and describe our measure of interest by willingness-to-pay.

prices to increase faster than wage levels, implying a premium for particular locations (Glaeser et al., 2001). This has led to a number of studies estimating the relevance of consumption opportunities to the desirability of places, with a focus on home-buyer decisions towards urban amenities. That is, by controlling for property-specific characteristics of a residential property such as the number of bedrooms, bathrooms, or whether the property has a back garden, the residual variation in the property value can be unpacked and related to the local availability of amenities or lifestyle opportunity. Under this approach, the desirability of urban environments has been shown to be factored into property values, and is broadly defined by the provision of place-specific assets and amenities that contribute to the allure of an urban area (Brueckner et al., 1999). Its importance, therefore, is intrinsically tied to population growth and development (Glaeser et al., 2001; Clark, 2004), as attractive places that elevate one's experience of an urban environment through concentrations of arts, leisure and retail have been shown to attract highly-skilled individuals (Florida, 2008). Clark (2004), for example, demonstrates that university graduates are more likely to locate to areas with high numbers of constructed amenities such as museums, libraries and leisure outlets. Oner (2017) pays particular attention to the role of retail as an urban amenity, regressing a Q-ratio – a ratio of the marginal price of a property to the marginal production cost – on variables reflecting accessibility to shopping destinations. In all, the study found a significant increase in the Q-ratio of 0.1 for every 1% increase in the accessibility to shops for city municipalities.

4.2.4 Measuring Retail Centre Attractiveness

In this paper, we follow methods of the second stream. Thus, we apply a hedonic framework to estimate the willingness-to-pay for retail centres. Given our focus on retail environments, business rates paid by commercial property such as high street shops provide an alternative, yet more suitable lens to explore hierarchies of retail centres than housing prices; while rent or housing prices are our idealised dataset, these are difficult to obtain, particularly at the national level. Under similar motivations to how urban economists proxy willingness-to-pay through residential housing, by controlling for property-level characteristics in *business rates* – the total floor area, the number of car parking spaces, the store type, for example – the remaining variation in a premise's business rate is convincingly explained by home-buyer desirability for a particular area, or, in our case, the retail centre. In the UK, non-domestic rates, or business rates, are a property-based tax levied on the estimated value of all non-residential properties such as shops, offices, warehouses and factories (Stuart and Miller, 2014). Business rates are determined using a rateable value for each non-domestic property. This is set by the Valuation Office Agency (VOA) who analyse rent evidence (rent and lease agreement details) alongside undertaking visual inspections of properties to ensure all evidence is considered fairly. VOA surveyors set rateable values to reflect features including: total floor area, business assets such as lifts, air conditioning and CCTV security systems, and changes in the local property market (VOA, 2014). A valuation begins by setting a common basic value per square metre for similar properties in the same area. This basic value is then adjusted to reflect the property's individual features. Each review of a property's valuation considers property-level characteristics and, most importantly, the buoyancy of the local property market. In this way, business rates are synchronized to local economic market conditions, reflecting the relative size and scale of retail economies (Astbury and Thurstain-Goodwin, 2014).

In our study, we label the estimated phenomena as retail centre willingness-to-pay (RWTP), which describes the price that home-buyers ascribe to the leisure and retail services offered by retail centres proximate to the property. In all, our findings are permissible because the residual variation in the business rate is attributed to local property market conditions (VOA, 2014), which themselves are influenced by home-buyer aspirations to reside in an environment that satisfies their wants and desires (Glaeser et al., 2001). By implication, this means the rateable value, once controlling for property-level characteristics, can be used to approximate RWTP for the retail centre whose catchment services the surrounding area. Under our conceptual approach, we can begin to unpack hierarchies of retail centres by undertaking a series of experiments on several econometric techniques to find a preferred specification that provides the most rigorous estimates of RWTP for retail centres across the case study of England and Wales.

4.3 Methodological Framework

Our approach to estimate retail centre willingness-to-pay (RWTP) relies on hedonic modelling (Rosen, 1974). This technique is typically used in the real estate literature to disentangle the price of a complex good as a function of the multiple intrinsic and extrinsic characteristics common to the property. In our case, a hedonic framework is applied to unpack the determinants of business rates for individual stores. By controlling for various property-level descriptors, a hedonic approach that uses a variable to represent each retail centre allows us to recover the implicit price for the retail and leisure opportunities provided by the retail centre. Practically speaking, this approach translates into a regression that explains the willingness-to-pay for receiving consumer amenities inside different retail centres. Once controlling for property-specific characteristics, the RWTP effect can be recovered for the location where stores are located because the business rate for each property involves setting a common basic value per square metre for *similar properties in the same area*, reflecting the performance, size and scale of local market conditions (Astbury and Thurstain-Goodwin, 2014).

To estimate the most robust empirical hedonic model specification, we compare several approaches, with a focus on recent contributions to the literature. To begin, we introduce a baseline spatial fixed effects model (Anselin and Arribas-Bel, 2013), which is expressed as:

$$\ln\, y_{ij} = \boldsymbol{x'_{ij}}\boldsymbol{\beta} + \sum_{j=1}^{J}\theta_j D_j + \epsilon_{ij} \tag{4.1}$$

where y_{ij}, the business rate for each store i in retail centre j, is log-transformed to alleviate the potential impact of heteroskedasticity; $\boldsymbol{x'_{ij}}$ is a $1 \times k$ vector of store-level variables in Table 4.3 and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficient to be estimated; D_j is the dummy variable for retail centre membership where $D_j = 1$ when $j = h$ for $i \in h$, 0 otherwise; and ϵ_{ij} is the model residual term, following an independent normal distribution $\mathcal{N}(0, \sigma_e^2)$. For model identification, the intercept is constrained to equal zero so that a separate RWTP effect θ_j can be estimated for each retail centre. From a non-technical standpoint, θ_j can be interpreted as the average willingness-to-pay (in log units) for stores to market their services in retail centre j. One might expect different retail centres offer varying degrees of utility such as access to particular socio-economic groups, amount of footfall, or the prestige of surrounding consumer amenities. Taking into account individual store characteristics, θ_j captures RWTP of retail centres.

Limitations exist associated with the fixed effect estimation strategy for the RWTP of retail centres. First, the estimator of θ_j, $\hat{\theta}_j$, would not be reliable and precise if the number of stores in retail centre j, n_j, is small. In addition, if different spatial processes operate at the property and retail centre scale, the conflation of unobservable influences will violate the independence of errors assumption through heteroscedastic and/or spatially-correlated error in the covariance structure (Dong and Wu, 2016). Multilevel models are approaches that allow variance between areas, and so they remedy these issues by treating the retail centre as part of the explanation for geographically varying outcomes (Owen et al., 2016). Instead of fitting a spatial fixed effect that assumes the relationship between the predictors and response holds as constant, multilevel models factor both spatial heterogeneity (differences) between areas and also dependencies (similarities) within them (Jones, 1991).

Put another way, this allows two stores located within the same retail centre to be more alike in their outcomes than would be expected given their individual characteristics alone. Correlation within boundaries are expected because stores are assumed to be affected by the same aggregate effects, also known as group dependence (Dong et al., 2015). Our second model thus requires a two-level hierarchical structure, an outcome variable measured at the lower level geography – individual stores – and a more aggregate spatial scale for the higher level – retail centres. We specify a random intercept multilevel model as,

$$\begin{aligned} \ln y_{ij} &= \boldsymbol{x'_{ij}}\boldsymbol{\beta} + u_j + \epsilon_{ij} \\ \text{var}(\epsilon_{ij}) &= \sigma_e^2;\ \text{var}(u_j) = \sigma_u^2. \end{aligned} \tag{4.2}$$

where u_j $(j = 1, 2, \ldots, J)$ measures the RWTP of the retail centre j, assumed to be independently distributed as $\mathcal{N}(0, \sigma_u^2)$. Under Eq 4.2, the dependency between stores in the same retail centre j is,

$$\text{cov}(y_{ij}, y_{ij}) = \text{cov}(u_j + \epsilon_{ij}, u_j + \epsilon_{ij}) = \sigma_u^2, \tag{4.3}$$

The random intercepts u_j are a linear combination of "fully-pooled" and "no-pooling" models. The fully-pooled model ignores heterogeneity by fitting a common intercept for all retail centre boundaries, whereas the no-pooling model, identical to the spatial FE, assumes a separate intercept for each retail centre. The multilevel model introduces the partial pooling, or shrinkage, of the RWTP effect towards the global intercept (Gelman and Hill, 2007). This is expressed as:

$$u_j = \tau_j u_j^{NP} + (1 - \tau_j) u_j^{FP}, \tag{4.4}$$

where u_j can be seen as a compromise between the "no-pooling" estimate u_j^{NP}, where each retail centre is assigned its own indicator variable, and the "fully-pooled" estimate u_j^{FP} which assumes a single intercept for all retail centres. This precision-weighted *compromise* is governed by the shrinkage factor τ_j (e.g. Goldstein, 2003),

$$\tau_j = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2/n_j)}. \tag{4.5}$$

where the weighting for τ_j is determined by the sample size in the $j-th$ retail centre (n_j) and the variation within (σ_e^2) and between (σ_u^2) groups (Goldstein, 2011). For example, when a retail centre boundaries contains a small number of stores n_j, the RWTP estimate is pulled towards the "fully-pooled" estimate. Similarly, when the boundary-level variance σ_u^2 is small – when the RWTP of retail centre boundaries is similar – estimates are more pooled towards the mean level than when σ_u^2 is large.

The use of a multilevel model in the estimation routines for constructing hierarchies of retail centres represents a novel application. Multilevel models have been used to produce league tables by inferring school effectiveness from individual pupil's educational attainment, but to our knowledge, have never been applied to explore hierarchies of retail centres. Moreover, whilst this area of social science has a rich history in the direct application of multilevel models (Goldstein, 2003), they rarely account for explicit spatial hierarchies in the empirical design. Thus, there has been emerging interest in incorporating spatial dependence into multilevel models (Dong and Harris, 2015). While we pursue a modelling objective similar to educational research by building a *league table* of retail centres, in the remainder of this section we develop an empirical strategy that accounts for potential spatial autocorrelation across the system of retail centres in space.

The model specified in Eq. 4.2 adopts a deterministic, 'container-driven' view of geographical space that contrasts with the reality that two retail centres located close together might be similar given their spatial proximity (Owen et al., 2016). In our case, we expect the RWTP effect induced by the retail centre at a particular location to be directly dependent on observed values at surrounding locations, with the intensity of this influence moderated by geographic proximity. This interaction is described by a simultaneous autoregressive (SAR) process. If the data generating process contains inherent spatial correlation, this may bias the estimated variance used for statistical inference. To account for this possibility in a spatially-explicit hierarchy, Dong and Harris (2015) distinguish between two kinds of spatial dependence: horizontal and vertical. The *horizontal* are the spatial dependencies between lower level units that are the traditional concern of spatial econometrics (Anselin, 1988), and the *vertical* are top-down group dependence due to regional effects. One potential problem is the vertical spatial dependence effect that causes the RWTP effect in nearby retail centres to be more similar than those further away. To account for this possibility, we specify a hierarchical spatial autoregressive (HSAR) model (Dong and Harris, 2015) that integrates SAR processes for the higher level residuals:

$$\begin{aligned} \ln y_{ij} &= x_{ij}\beta_k + \theta_j + \epsilon_{ij}, \\ \theta_j &= \lambda M_j \boldsymbol{\theta} + u_j, \end{aligned} \tag{4.6}$$

where M is an $J \times J$ spatial weights matrix that captures the interaction structure of stores by assigning non-zero weight $M_{ij} \neq 0$ to pairs of observations assumed to be spatial neighbours, zero otherwise. M_j is the j-th row of M. Given the spatial characteristics of the data points, we define neighbours using an exponential decay function with the distance bandwidth d set to five kilometres[2]. Following convention, M is row-standardized so that each row sums to unity $\sum_j M_{ij} = 1$. The parameter λ quantifies the correlation of

[2]Spatial connectivity at the retail centre level is specified as

$$M_{ij} = \begin{cases} 1, exp(-(d_{ij}^2)/d^2), & \text{if } d_{ij} \leq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

where d_{ij} is the Euclidean distance between retail centres and d is the fixed distance bandwidth. A semivariogram was used as an exploratory tool for determining the distance at which the spatial dependence between business rates between retail centres became negligible (see Figure 4.5).

RWTP, with higher values for λ leading to spatial covariance that dissipates slower for a higher order of neighbours. The reduced form of $\boldsymbol{\theta}$ in Eq. 4.6 is

$$\boldsymbol{\theta} = (I_J - \lambda M)^{-1}\boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, I_J\sigma_u^2) \tag{4.8}$$

where the spatial filter $(I_J - \lambda M)^{-1}$ captures any vertical spatial dependence in the RWTP effect θ_j. A "Leontief expansion" of the matrix inverse expands to $(I_J - \lambda M)^{-1} = I + \lambda M + \lambda^2 M^2 + \lambda^3 M^3 + \cdots$ and demonstrates spatial feedback when an increasing order of neighbours creates bands of ever larger reach around each location, relating every retail centre to every other one (Anselin, 2003).

A different, but related, model we specify next is a hierarchical spatial error model (HSE) which is similar to Equation 4.6, except that we specify a spatially-autocorrelated error term in η:

$$\begin{aligned} \ln y_{ij} &= x_{ij}\beta_k + \theta_j + \epsilon_{ij}, \\ \theta_j &= u_j + \lambda M\boldsymbol{\eta} \end{aligned} \tag{4.9}$$

A final methodological consideration relates to LeSage (2014)'s empirical question as to whether the spatial process under study is *global* or *local*. The covariance structure induced by the HSAR model is global, as the spatial process relates every retail centre to each other. A spatial moving average (SMA) process, on the other hand, considers only first- and second-order neighbours, beyond which the spatial covariance is zero (Anselin, 2003):

$$\theta_j = \gamma M\boldsymbol{\theta} + u_j. \tag{4.10}$$

The data generating process of Eq. 4.10 collapses to the reduced form

$$\theta_j = (I_J + \gamma M)\boldsymbol{u}, \tag{4.11}$$

where everything holds as in Eq. 4.6, except that we introduce the SMA parameter γ. Unlike Eq. 4.8, because $(I_J + \gamma M)$ is not inverted in the SMA specification, there is only local range for the induced spatial covariance. This approach is intuitive as there may only be local interaction across a 'neighbourhood' of different retail centre boundaries, as opposed to interaction across the entire system of the national extent.

While the standard multilevel model is estimated using maximum likelihood estimation, the spatial models are estimated using a Markov Chain Monte Carlo (MCMC) simulation technique whose stationary distribution constructs a target probability distribution for the parameters. MCMC simulations are typically the only feasible approach for fitting spatial models that introduce the complexities of place relatedness into the variance-covariance structure (Lesage, 1997). Under these motivations, conditional Gibbs samplers are derived for the HSAR (Dong and Harris, 2015) and HSE and HSMA (Wolf et al., 2020) models to obtain posterior samples for each parameter. This way, the joint density for the parameters are broken into univariate conditional probabilities where every successive parameter draw is conditioned on the draw for the previous parameter value (Geman and Geman, 1984). Not only is this sampling technique computationally efficient, the draws from the parameter space $\{\beta, \sigma_e^2, \sigma_u^2, \lambda\}$ accumulate to an entire distribution for each parameter. In our case, we summarise each parameter estimate by the median value across the distribution, but also with interval calculations. Each sampling chain is simulated for 10,000 iterations, with the first 5,000 draws discarded as "burn in" to allow the posterior distributions for each parameter to converge. In addition, we assess the serial autocorrelation for the posterior draws by examining the effective number of independent samples. As in time series analysis, we evaluate this because autocorrelation can often understate estimates of the variance in correlated sequences. A final methodological note, the same weakly informative prior distributions were assigned to the model parameters in each model[3]. Further details on the technical implementation of the spatial models for the HSAR are found in Dong and Harris (2015), and the HSE and HSMA in Wolf et al. (2020).

Despite our empirical strategy becoming increasingly sophisticated, a commonality between each model is that we obtain a free measure of uncertainty alongside each estimate of the willingness-to-pay for a particular retail centre, θ_j. Uncertainty is expressed in the estimates for the confidence intervals of the spatial FE and MLM, and Bayesian credible intervals for the HSAR, HSE and HSMA models. As point estimates for θ_j represent an absolute ranking, overlapping interval estimates for each retail centre imply confidence/credibility regions that change the rank ordered estimate of centres in the hierarchy. Where the density bands of the confidence or credible intervals become less disjoint, there is increased uncertainty in the disambiguation between ranks of a given set of retail centres. Uncertainty measurements are desirable in cases where retail centres contain a small number of stores n_j. Returning to Eq. 4.4, as this carries implications for the calculated u_j, an uncertainty estimation is valuable to ascertaining a measure of trust in the rankings of retail centres.

4.4 Data

Our point of departure for the proposed methodological approach is a geographical data set sorted into a hierarchical structure consisting of units grouped at two different levels. The points in our lower level geography represent 355,076 individual high street stores

[3] The following conjugate priors are chosen:

$$
\begin{aligned}
P(\beta) &\propto \mathcal{N}(0, 100) \\
P(\sigma_e^2) &\propto InverseGamma(0.01, 0.01) \\
P(\sigma_u^2) &\propto InverseGamma(0.01, 0.01) \\
P(\lambda) &\propto Uniform(-1, 1)
\end{aligned}
$$

across England and Wales that are located inside retail centre boundaries. This includes franchised chains such as fast-food outlets, supermarkets and clothing stores – McDonald's, Tesco and Primark, for example – but also independent retailers with more local scope. The data was collected by a large pool of surveying teams from the Local Data Company (LDC) in 2015, and includes various descriptors for each property such as retail function and occupancy status. The most important characteristic of the data is that commercial addresses in the LDC database are matched to addresses in the VOA 2010 rating list (VOA, 2018). This affords us a business rate valuation for every non-domestic premise, allowing us to unlock a rich, unique and highly granular dataset that provides a new and alternative lens through which to explore the implicit value describing the willingness-to-pay of an area. In all, for every store we have store-level variables that offer a rich description of the premise's physical condition. This includes data collected by VOA surveyors on the date of assessment such as the total floor area, the amount of rooms in the premise and the number of car parking spaces, but also data collected by the LDC that categorises the business's function[4]. A full description of the variables that enter our design matrix is tabulated in Table 4.3.

Yet, there are limitations to the VOA rating list that introduce error, especially given the primary purpose of the list is not intended for data analysis. The most notable limitation is what Astbury and Thurstain-Goodwin (2014) describe as the regional difference in data collection techniques that affect the extent to which the rateable value reflect the market tone of a particular area may lead to over- and under-predictions of the business rate assessed for the premise. Moreover, whilst the rating list was released in 2010, the rateable values set are actually conditioned on the 2008 market climate. Given the UK economy underwent the shock of an economic crisis during this period, a time characterised by fragile consumer confidence, a decline in household disposable incomes, and rising shop

[4]LDC premise types were recoded in accordance to VOA Special Categories outlined in Rhodes and Brien (2017).

vacancy on the high street (BIS, 2011), it is likely the overall market tone has been over- and under-valued across retail centres for England and Wales. Despite these considerations, the VOA ratings list provides highly granular and geographically accurate access to data reflecting local market economic conditions for the national extent.

The retail centre is the observational unit from which we obtain home-buyer willingness-to-pay estimates. Our higher level units are represented by 2,951 exogenously-determined retail centres across England and Wales. Conceptually, the retail centre is an appropriate choice for this purpose because they are both drivers of local economic performance and reflect the wider economic health and social well-being of the urban environment (BIS, 2011). Moreover, while they are often viewed as hubs for retail activity, they also exhibit a multitude of heterogeneous uses, including services, offices, residential and public buildings (Teller and Elms, 2012). The boundaries used in the present study were produced by Pavlis et al. (2018) as a successor to boundaries developed by Thurstain-Goodwin and Unwin (2000) for the Department for Communities and Local Government (DCLG) in 2004, with the exception that they were intended to move away from a definition of town centre locations of employment to functional spaces delineated for retail. While the resulting retail centres may not perfectly align with those designated in governmental planning policy, they provide a consistent method for comparing retail centres nationally. In all, these boundaries are our higher level geographical unit, and represent the functional economic market area of the retail centre. The resulting spatial hierarchical structure of the data is illustrated in Figure 4.1 through the example of Liverpool.

4.5 Empirical Findings

In this section, we develop a discussion of our empirical findings in two main directions: first, we step through each of the modelling approaches, building intuition towards our

Figure 4.1: Lower level store premises (black dots) nested into higher level retail centre boundaries (in red) for Merseyside, UK.

preferred specification; and second, we introduce a validation exercise to evaluate whether variation in the estimated retail centre willingness-to-pay (RWTP) effect can be attributed to characteristics that are generally associated with attractive areas.

4.5.1 Model Validation

Our point of departure is a discussion of the results provided in Table 4.1 for the proposed methodology[5]. Thus, before exploring the subtleties of the multilevel specifications, we firstly step through a description of the parameter estimates for the store-level explanatory characteristics. To do this, we use the classical multilevel model (shown by column (2) in Table 4.1) as a baseline, but note the estimates are generally consistent across each model. Overall, the estimates for the store-level covariates that enter our design matrix are fairly intuitive and of the expected signs for all models. For example, every additional room in the premise increases the rateable value by 7%, which is consistent with the VOA's mandate to adjust the rateable value by property-level characteristics (VOA, 2014). This is also reflected in the number of car parking spaces, where each additional ten spaces increases the rateable value by 1.3%. Somewhat surprisingly, increasing the total floor area by $1,000m^2$ only seemed to increase the rateable value by 2.3%, but given we control for different store sizes latently with the store category variables, this is somewhat expected. On the whole, the store type categorisations are consistent with conventional wisdom. The rateable value for premises such as takeaway food outlets, for example, are generally 20.2% less than the reference category, showrooms. This makes sense because the locations of takeaway outlets are generally linked to geographical inequalities in health outcomes (Daras et al., 2018), which are simultaneously related to environments that are considered less desirable. On the other end, the rateable value for hypermarket stores (with a gross floor area over $2,500m^2$)

[5]Potential problems of multicollinearity were assessed using variance inflation factor (VIF) scores for each predictor variable in the spatial FE model. VIF scores revealed no evidence of such problems, with scores of about 3.0 leading us to continue with our inferential exercise.

are three times greater, which is expected given the number of business assets such as lifts, warehouse machinery and CCTV security systems common to large supermarket stores.

We next address model selection by means of goodness-of-fit tests. In each case, every model in Table 4.1 had a highly similar Root Mean Square Error (RMSE) and log-likelihood value. While the R^2 of the spatial FE model (67.6%) is marginally higher than the multilevel model(s) (67.1% to 67.2%), the spatial FE fits a parameter for $J = 2,951$ retail centres, which contrasts with the regularisation introduced by hierarchical pooled effects in the multilevel models for smaller groups. In other words, not only does the spatial FE likely overfit, but the estimates and standard errors of the retail centre fixed effect will be noisier in places with a smaller number of properties. In our case this is pertinent because the minimum number of stores across retail centre boundaries is two. For this reason, we motivate our preferred specification as the multilevel model(s). However, as the performance of each multilevel specification is comparable on goodness-of-fit grounds, we undertake further examination of the substantive effects in the RWTP estimate later on.

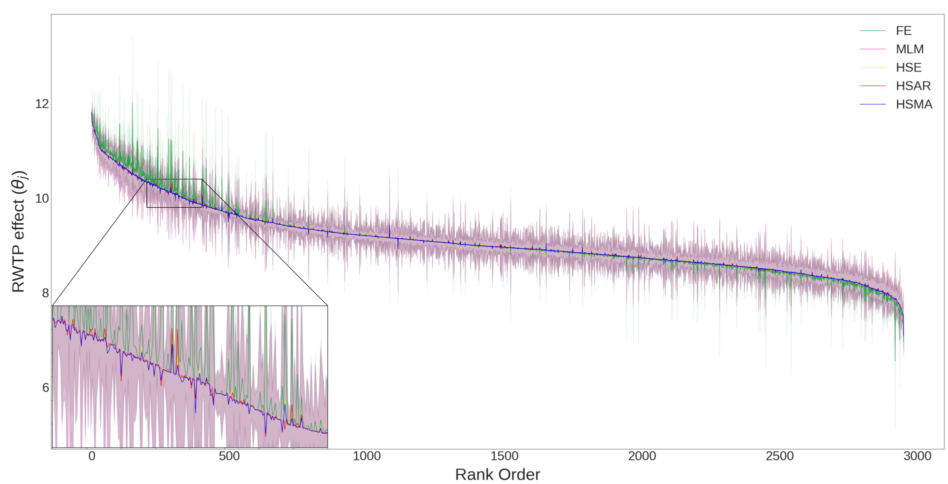


Figure 4.2: Rank ordered estimates for retail centre willingness-to-pay (RWTP). *Note*: 95% confidence and credible intervals are shaded either side of the point estimate.

Table 4.1: Regression coefficients estimates for estimated models.

	Dependent variable:				
	ln Business Rate				
	FE	*MLM*	*HSAR*	*HSE*	*HSMA*
	(1)	(2)	(3)	(4)	(5)
(Intercept)	NA	9.088***	9.064***	9.067***	9.071***
	NA	(0.015)	(0.018)	(0.026)	(0.027)
Structural Characteristics					
No. Rooms	0.070***	0.069***	0.069***	0.069***	0.069***
	(0.000)	(0.0003)	(0.000)	(0.000)	(0.000)
Floor Area	0.023***	0.023***	0.023***	0.000***	0.023***
	(0.0003)	(0.0003)	(0.0003)	(0.0000)	(0.0003)
Car Parking Spaces	0.013***	0.013***	0.013***	0.013***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Store Typology					
Banks and Other A2 Uses	0.048***	0.046***	0.048***	0.046***	0.046***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Factory Shops	0.060***	0.057***	0.058***	0.057***	0.057***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)liv
Food Stores	0.066***	0.061***	0.061***	0.061***	0.061***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
Hairdressing/Beauty Salon	−0.356***	−0.360***	−0.360***	−0.360***	−0.360***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Hypermarket/Superstore	3.037***	3.053***	3.054***	3.053***	3.054***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
Large Food Stores	1.556***	1.568***	1.569***	1.569***	1.569***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
Large Retail Shops	2.770***	2.770***	2.766***	2.764***	2.760***
	(0.427)	(0.427)	(0.425)	(0.429)	(0.427)
Non-Retail	−0.385***	−0.388***	−0.388***	−0.388***	−0.388***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Other	−0.312***	−0.316***	−0.316***	−0.316***	−0.316***
	(0.0125)	(0.012)	(0.013)	(0.013)	(0.013)
Pharmacies	−0.197***	−0.200***	−0.200***	−0.200***	−0.200***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Post Offices	0.001	−0.003	−0.002	−0.003	−0.002
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Restaurants and Bars	−0.051***	−0.051***	−0.051***	−0.051***	−0.051***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
Retail Shops	0.108***	0.106***	0.106***	0.106***	0.106***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Takeaway Food Outlet	−0.199***	−0.202***	−0.202***	−0.202***	−0.202***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Variance Components					
σ_e^2	0.541	0.541	0.541	0.541	0.541
	(0.0001)	(0.0001)	(0.001)	(0.001)	(0.001)
σ_u^2	NA	0.503	0.484	0.477	0.492
	NA	(0.0001)	(0.014)	(0.014)	(0.014)
λ	NA	NA	0.232***	0.230***	0.189***
	NA	NA	(0.026)	(0.023)	(0.020)
RMSE	0.732	0.733	0.747	0.733	0.732
pseudo-R^2	0.676	0.671	0.671	0.671	0.671
Log Likelihood	-395,353.5	−402,548.3	-395,478.8	-395,481.9	-395,479.2
AIC	796,644.9	805,136.7	790,993.7	790,999.7	790,994.4

Note: p-values for Bayesian models correspond to credibility intervals crossing zero. *p<0.1; **p<0.05; ***p<0.01

A comparison of the rank ordered estimates for the RWTP effect θ_j are visualised for each model in Figure 4.2. The figure reflects our rankings of point estimates for RWTP, along with a measure of uncertainty shown by the 95% confidence (FE and MLM) and credible (HSAR, HSE, HSMA) intervals. If any of the confidence/credible density bands for any two models overlap, the two estimated ranks are not distinct. The rankings, 1 to 2,951, are presented on the x-axis, while the y-axis displays the estimated RWTP value in log units. Additionally, we include a zoomed inset to highlight movement in the estimated RWTP value, which is zoomed at a window that displays the most variability in the estimated scores between each model. Taking a closer look, it appears the movement for the RWTP estimate relative to the spatial FE model, marked by the green line, is not uniform. In the upper and lower tails, for example, there is systematic variation in the parameter estimates between the spatial FE and estimates of the multilevel models. This suggests the point estimates for RWTP values deviate widely from the multilevel models for the *most* and *least* desirable retail centre boundaries, with there being little systematic variation in-between. At a general level, the figure reproduces a classical result, as the estimates of the multilevel model demonstrate hierarchical pooled effects, that is, *shrinkage* towards the global intercept. Here, the estimates exhibit improved precision, which contrasts with the higher magnitude of uncertainty in the spatial FE estimates, as shown by the more extreme and noisy estimates in the upper and lower tails of the figure. Shrinkage effects can be seen clearer in Figure 4.3, where we sample nine retail centres from our rankings to demonstrate movement in the RWTP estimates by expanding the point estimates horizontally along a 2-D axis for each model. In the case of Meridian Leisure Park, Leicester, for example, the FE estimate is shrunk from 10.42 $\pm$.73 to 9.72 $\pm$.51 in the MLM. In real terms, this reflects a change in magnitude from £33,523.43 to £16,647.24 when we exponentiate from log units. Interestingly, what is also observable for this retail centre is what Wolf et al. (2020) describe as "spatially-local shrinkage", where spillovers from the jth adjacent retail centres cause *growth* in the spatial multilevel estimates towards the mean of neighbouring retail centres from 9.72 ±.51 to 9.82 ±.51 under the HSAR model. While none of the interval estimates become disjoint for each retail centre in Figure 4.3, the findings from the spatial models *suggest* the RWTP estimate is moderated by shrinkage towards the values of neighbouring retail centres.

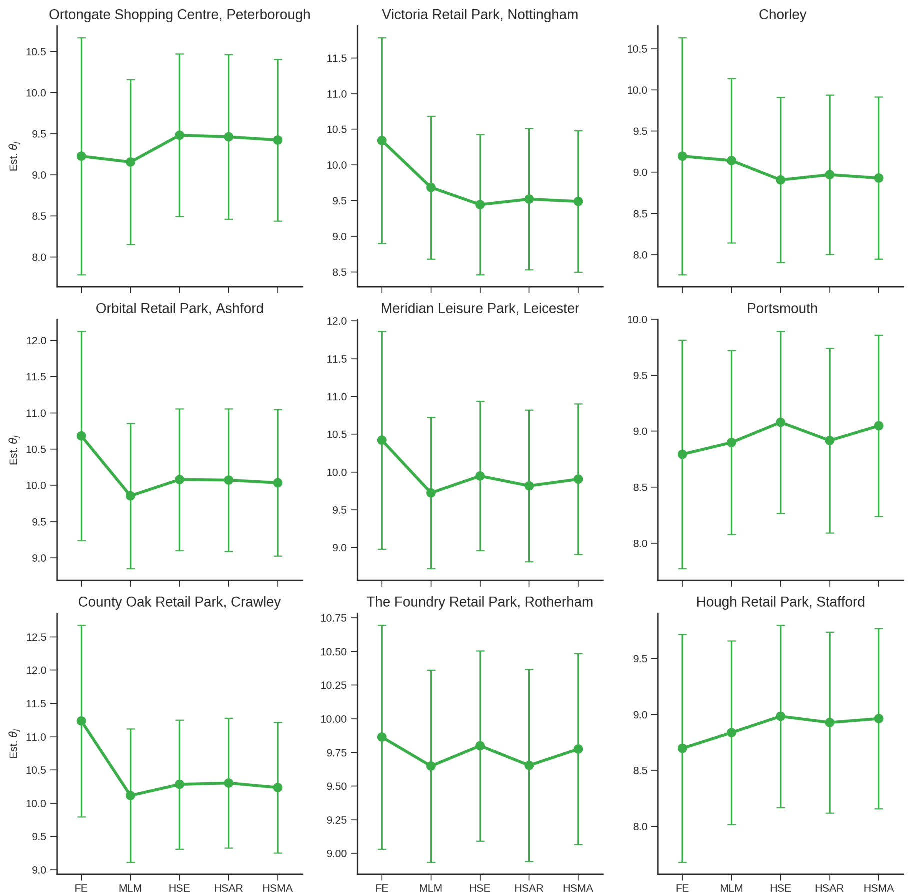


Figure 4.3: Model estimates for nine retail centre boundaries sampled from the hierarchy of rankings.

Having discussed our rankings, we now build intuition towards our preferred specification for the RWTP estimate, which we begin by turning attention to the within-boundary (σ_e^2) and between-boundary (σ_u^2) variance components. By combining these measures, we calculate the variance partitioning coefficient (VPC) for the MLM which measures the proportion of variance explained by the hierarchical structure ($\frac{\sigma_u^2}{\sigma_u^2+\sigma_e}$). This measure outlines the correlation between stores within the same retail centre, and is required to ascertain the percentage of variation explained by the retail centre differences for store i in retail centre j (Browne et al., 2005). The VPC statistic reveals a value of 0.482, meaning 48.2% of the variance in the response is explained by the retail centre geography. This VPC value motivates the empirical decision to take our multilevel models as the preferred specification(s) over the FE model, with these models able to flexibly accommodate the covariance structure induced by the grouping of stores by retail centre boundary. Our search for a preferred specification continues by evaluating potential spatial dependence in the RWTP effect u_j estimated by the MLM. Given the MLM assumes RWTP values to be independent of each other, we follow Dong et al. (2015) and use a Moran's I to test whether the estimates for RWTP are spatially dependent. A Moran's I statistic for u_j premised on the spatial weights matrix M for the retail centre polygons returns a coefficient value of 0.174 (p-value equal to < 0.001). This illustrates positive spatial autocorrelation for the estimated RWTP values, which motivates using the spatial models given the core model assumption of independence for u_j across retail centres does not hold.

We subsequently turn direct attention to the spatial multilevel models. Given our hierarchical approach is fully Bayesian, trace plots are required to monitor the convergence of each parameter to the target distribution (see Figure 4.6). In each case the parameters were assessed to have converged. Moreover, there was no serial autocorrelation identified in the stationary Markov chain for each parameter. The first substantive difference we observe is that not accounting for spatial dependence leads the MLM to marginally overestimate the retail centre boundary variance σ_u^2 relative to the spatial models; σ_u^2 can be understood as the average variation of RWTP values across the retail centres. Here, σ_u^2 falls from 0.503 in the MLM to 0.484, 0.477 and 0.492 in the HSAR, HSE and HSMA models, respectively. We also recover evidence of a significant spatial autoregressive parameter

λ, which is indicative of spatial spillovers effects of RWTP values between neighbouring retail centres. This is recovered as λ is distinct from zero at the 95% credible interval. Interestingly, the density of the covariance structure seems to impact the estimate for λ. The HSMA model, with a sparse covariance structure that is restricted to first- and second-order neighbours, estimates a λ value of 0.189. On the other hand, models with a denser covariance structure such as the HSAR and HSE estimate highly similar values of 0.232 and 0.230. Each of these estimates indicate spatial interaction effects among retail centre boundaries.

To aid the visualisation of spatial patterning, we illustrate the case of Liverpool in Figure 4.4 with assistance of *legendgrams* that show the distribution of RWTP values across all 2,951 retail centres, and is colour coded using $k = 8$ break points classified using Fisher-Jenks optimization (Jenks, 1967). Each cell highlights a selected retail centre in red, with the corresponding RWTP estimate shown by the vertical bar stemming from the x-axis of the legendgram, with 95% confidence and credible intervals shaded either side to highlight uncertainty in the estimate. From left to right, the columns identify the RWTP estimate for the FE, MLM, HSAR and HSMA models. From a first reading, the spatial patterning in the figure seems to reveal a fragmented picture of vitality and decline, with less desirable retail centres observed in the immediate hinterland of the prospering regional centre (identifiable by the large red polygon in the first row). Overall, from this reading of the figure, we are able to discern spatial hierarchies that possibly fragment Merseyside's functional market area, with certain retail centres eliciting a higher willingness-to-pay than neighbouring centres.

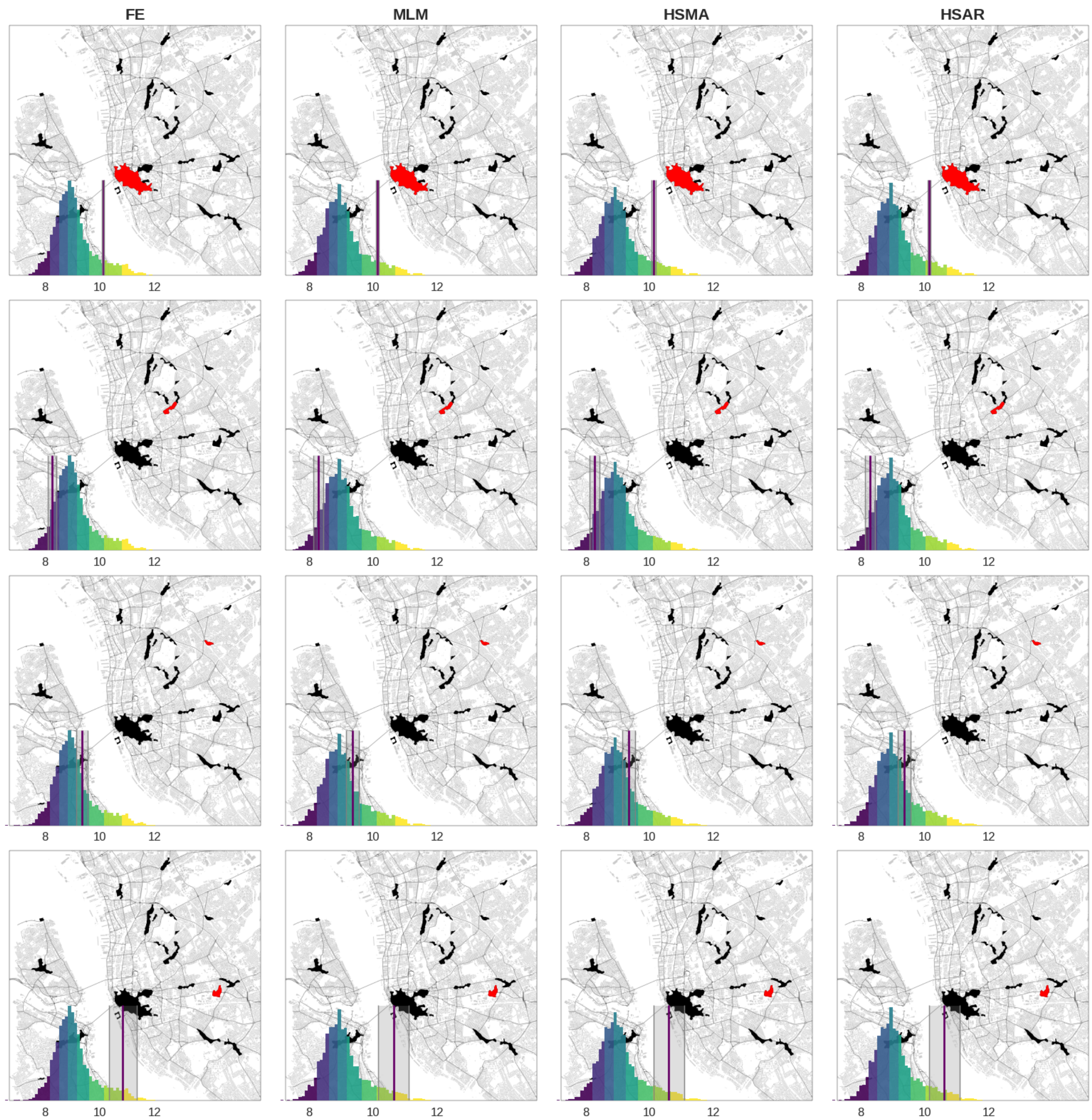


Figure 4.4: Retail centre willingness-to-pay (RWTP) values in log units for Merseyside, UK. *Note*: purple vertical bar identifies the value of the retail centre highlighted in red, with 95% confidence/credibility intervals shaded in grey either side.

4.5.2 Technical Validation

After motivating our preferred methodological approach, we undertake a validation exercise to evaluate whether the estimated RWTP effect θ_j for each retail centre in the HSAR model responds to characteristics that are generally identifiable for prospering and thriving areas. Here, we regress θ_j on a selection of variables using ordinary least squares to: firstly, assess whether any of the variation in the estimated RWTP values can be attributed to variation in the selected explanatory variables; and secondly, to quantify the strength of relationship, if any, between the response and explanatory features. Principal attention is paid to the 2011 census Workplace Zone (WZ) population characteristics (Mitchell, 2014) that represent individuals working in the retail centre. As commuter patterns change the spatial distribution of the working population, which holds when the bulk of economic activity occurs during "traditional" office hours (Mitchell, 2014), WZ statistics are preferable because they describe the day-time working population who commute to their places of work inside the retail centre. The WZ variables we use include: the percentage of people who report their general health as "Good" or above; the percentage of individuals with no qualifications; the percentage of home-owners; the percentage of workers enrolled in higher managerial occupations; and the percentage of individuals in full-time employment. Other variables we consider include: the vacancy rate of stores in the retail centre calculated from the LDC database[6]; a raw count of stores from the LDC database; the amount of urban green space (m^2) (Daras et al., 2018); logged median housing values for the 2015 rolling year (Land Registry, 2016); and finally binary variables for regions in England and Wales that reflect Nomenclature of Territorial Units for Statistics (NUTS) subdivisions – North West, London, West Midlands and Wales, for example. In each case, the variables are spatially joined[7] from WZ statistical units to the retail centre boundary polygons.

[6] Vacancy rates are defined as the proportion of all available retail units that are vacant or unoccupied.

[7] As there is only partial overlap between the retail centres and WZ polygons, the resulting WZ statistics are aggregated by the spatially weighted mean value for the intersecting WZ geometries when joined to each retail centre polygon.

The findings are displayed in Table 4.2. Generally, they are consistent with expectations, although there are deviations from conventional wisdom. For WZ characteristics, an increase to the number of individuals with "Good" health (or above) by one percent increases the RWTP value by 3.9%. Similarly, an increase in the number of people with no qualifications by one percent decreases the value by 2.9%. Surprisingly, an increase in the number of workers in higher managerial occupations by one percent decreases the RWTP of the retail centre by 4.9%. At first glance this result appears counter-intuitive, but managerial workers are more likely to work in financial districts characterised by mostly office space, which are not necessarily perceived as desirable in the same way consumer amenities such as leisure plazas and urban green spaces are.

Next, we consider retail centre boundary characteristics. For every additional one hundred stores in the retail centre, the RWTP value increases by 4.30%, which is implicit of patrons valuing a large number of available retail destinations. Similarly, as the vacancy rate increases by 1%, the RWTP of the area decreases by 2.2%. Again, this is consistent with expectations that a large number of vacant units deteriorate the vibrancy of the streetscape by revealing signs of decay. On the other hand, the availability of urban green space was not a significant determinant. For the regional indicators, relative to the East Midlands reference category, we recover some examples of regional inequality. While retail centres in the East of England are estimated as having the highest RWTP value (28.9%), there is a clear disparity in the estimated values for North West England (9.5%), South West England (4.1%) and to a less extent Yorkshire and The Humber (13.5%) when compared to South East England (20.3%) and London (20.3%). These inequalities are broadly consistent with regional variations in wealth across England and Wales (Rowlingson and McKay, 2011). In all, the validation exercise demonstrates a relationship between RWTP values and socio-economic characteristics that is consistent with conventional wisdom. Although not conclusive, the coefficients of our estimates suggest a decline in RWTP is related to urban environments that with poorer social and community well-being. This *begins* to address a key gap in the evidence linking retail centre outcomes to characteristics of the urban environment that is identified by The Department of Business, Innovation and Skills for England and Wales (BIS, 2011).

Table 4.2: OLS regression results for validation exercise.

	Dependent variable:
	HSAR θ_j
	(1)
(Intercept)	2.684***
	(0.485)
Workplace Zone Characteristics	
Good Health (%)	0.039***
	(0.002)
No Qualifications (%)	−0.029***
	(0.005)
Tenure Owned (%)	0.0003
	(0.001)
Higher Managerial Occupations (%)	−0.049***
	(0.003)
Full-time work (%)	0.020***
	(0.001)
Retail Centre Boundary Characteristics	
Vacancy Rate	−0.017***
	(0.002)
Store Count	0.024***
	(0.004)
Urban Green Space	−0.002
	(0.018)
ln Median House Price 2015	0.243***
	(0.039)
Regional Indicators	
East of England	0.289***
	(0.056)
London	0.203***
	(0.061)
North East England	0.095
	(0.067)
North West England	0.023
	(0.053)
South East England	0.203***
	(0.052)
South West England	0.041
	(0.054)
Wales	−0.066
	(0.068)
West Midlands	0.229***
	(0.057)
Yorkshire and The Humber	0.135**
	(0.054)
Observations	2,951
RMSE	0.555
Adjusted R^2	0.352
Note: regions reference category is East Midlands.	*p<0.1; **p<0.05; ***$town$p<0.01

4.6 Conclusion

The depth and breadth of leisure and retail opportunity is increasingly linked to the desirability of places to live (Glaeser et al., 2001). As the quality of urban environments cannot be qualified by a natural unit of analysis, the willingness-to-pay to receive an amenity-rich environment has often been explored through the lens of the residential housing market. The groundings of this paper were motivated by similar hedonic analyses, except that we used business rates for commercial properties alongside a non-trivial methodological framework to estimate retail centre willingness-to-pay (RWTP), of which we provide a detailed exposition for reproducing the analysis. Similar to approaches that analyse housing prices, by controlling for property-level characteristics such as the total floor area, car parking spaces, and store type, the remaining variation in the business rate was attributed to the willingness-to-pay of the retail centre. This was possible because business rates approximate local market conditions, as rateable values are set by estimating a basic cost per square metre which is adjusted to reflect similar properties in the same area (VOA, 2014). Despite our empirical motivations, particular attention to how the RWTP estimates interface with the unique geographic behavioural characteristics of the UK retail landscape was required. Due to restructuring of the traditional brick-and-mortar retailer landscape through growth in electronic retailing our study required particular attention to the nuances of UK retail spaces. It is often argued that growth in online retailing is forecast by its deleterious effects that cause physical shopping opportunity to be substituted online (Doherty and Ellis-Chadwick, 2010). Despite these concerns, online retail has recently been linked to complementarity and modification processes. These processes blend traditional retail with e-commerce through integration of technologies such as ‘click and collect’ points that operate as points of delivery for internet sales (Singleton et al., 2016). Thus, through the market system of using business rates, the RWTP estimates relate to how much the behaviour of consumers value a given retail area. Amongst the context of behavioural patterns, this allowed us to unpack hierarchies of retail spaces. These spaces are an underlying driver to the sustainability of built environments and so, by implication, reveal geographical patterns in urban growth and development.

Multilevel models have a rich history in the educational sciences literature for building league tables of school performance (Goldstein, 2003). We used similar motivations to build a ranking of retail centres, except that unlike previous studies, we allowed for possible spatial autocorrelation that operates on the basis of geographical proximity. This is because the RWTP effect per retail centre is likely to covary based on spatial proximity. Under these motivations, and by revamping the traditional focus of multilevel modelling techniques, we were able to derive retail centre estimates of RWTP. A particular focus on retail centres, our geography of choice, was because they have been argued as a moderating influence on urban hierarchies (Dennis et al., 2002). Yet, there is a limited availability of national data for measuring the economic and social value of retail centres, with a presumptive attitude in UK policy circles that the impacts of policy instruments such as the Town Centres First approach are "instinctively positive" (BIS, 2011). In producing ranked estimates, we remedied these uncertainties by building quantifiable evidence to directly observe disparities in RWTP across networks of retail centres. More concretely, the derived scores allow an understanding of a particular retail centres position within a network of centres; this can be used as a proxy of economic health and an indicator of the gravity that particular retail centre catchments pull on consumers in the area. From this retail practitioners might be able to use the derived scores as proxies for footfall generation which would allow them to deduce consumer appeal of particular centres. Knowledge of such characteristics might be used in decision-making processes such as determining investment and divestment outcomes or the rationalisation of store portfolios, for example. At a general level, our findings also provide a platform for researchers to build upon. The applied methodology provides a blueprint for constructing hierarchies of retail centres that is replicable and generalisable to similar contexts, conditional on data availability. Not only this, to our knowledge, the study is a first of its kind that to build indicators that describe hierarchies of retail centres across a national extent, with previous studies typically limited to smaller case study areas. Finally, a core and intentional contribution of the paper is the potential for exploration of hypotheses in retail geography that were previously unavailable due to the absence of statistical data on retail centres.

To conclude this article, we illustrate elaborations to consider for future research. One refinement involves the addition of further attributes at the store or retail centre level to be specified into the modelling approach. This might involve undertaking visual, in-person surveys for small case study areas to collect image attributes identified in Gomes and Paula (2017) such as parking security, atmosphere perception, or mix and quality of stores within the retail centre boundary, for example. Due to the practicality concerns of obtaining these highly granular measures in the present study, this direction would reduce the number of retail centres the approach can return RWTP estimates for. However, the benefit is that it would allow an estimation of the willingness-to-pay for highly granular measures that describe image-based attributes of attractive shopping environments. A final remark, the advantage of the applied methodology is that it can be redeployed in the future to generate timely updates. This is possible because the VOA continue to reassess the rateable values of non-domestic properties according to a five year revaluation cycle (VOA, 2014). Conditional on the VOA continuing to release their ratings list as an open data product, the area estimates of RWTP are updatable over time. Future research might develop retail centre rankings into a longitudinal data product that allows an exploration into the temporal characteristics of RWTP, and how successive five year windows alter the rank-ordered positions of retail centres.

4.7 Appendix

4.7.1 Appendix A: Variable Description

Variable	Description	Source	Mean	Std. Dev	Unit
Dependent Variable					
Business Rate	Rateable value taxed on the business property.	VOA	100,632.9	976,771.5	Pounds
Structural Characteristics					
Floor Area	Total floor area for the property (0,000s).	VOA	0.658	6.155	m^2
No. Rooms	Number of surveyable rooms.	VOA	6.577	4.010	Count
Parking	Number of car parking spaces (00s).	VOA	0.099	2.053	Count
Store Typology					
Banks and Other A2 Uses	1 for A2 uses, 0 otherwise.	LDC	0.07	0.25	Binary
Factory Shops	1 for factory shop, 0 otherwise.	LDC	0.004	0.07	Binary
Food Stores	1 for food store ($< 750m^2$), 0 otherwise.	LDC	0.04	0.20	Binary
Hairdressing/Beauty Salon	1 for salon, 0 otherwise.	LDC	0.13	0.33	Binary
Hypermarket/Superstore	1 for superstore ($> 2500m^2$), 0 otherwise.	LDC	0.03	0.16	Binary
Large Food Stores	1 for large food store ($> 750m^2$), 0 otherwise.	LDC	0.04	0.20	Binary
Large Retail Shops	1 for large shops ($> 1850m^2$), 0 otherwise.	LDC	0.04	0.20	Binary
Non-Retail	1 for non-retail, 0 otherwise.	LDC	0.05	0.21	Binary
Other	1 for 'other' premises, 0 otherwise.	LDC	0.01	0.12	Binary
Pharmacies	1 for pharmacy, 0 otherwise.	LDC	0.004	0.06	Binary
Post Offices	1 for post office, 0 otherwise.	LDC	0.01	0.10	Binary
Restaurants and Bars	1 for restaurant or bar, 0 otherwise.	LDC	0.07	0.25	Binary
Retail Shops	1 for high-street retail store, 0 otherwise.	LDC	0.40	0.49	Binary
Takeaway Food Outlet	1 for takeaway outlet, 0 otherwise.	LDC	0.11	0.32	Binary
Showrooms	1 for showroom, 0 otherwise.	LDC	0.03	0.18	Binary

Table 4.3: Variable description for property-level characteristics.

4.7.2 Appendix B: Semivariogram for Business Rate Semivariance

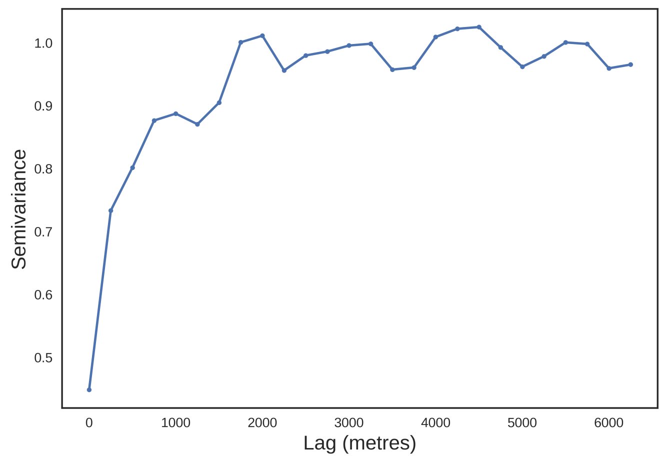


Figure 4.5: Semivariogram demonstrating the tendency for retail centres close together in space to exhibit higher correlations for business rates than those further apart.

4.7.3 Appendix C: Trace Plot for Markov Chain Monte Carlo Draws

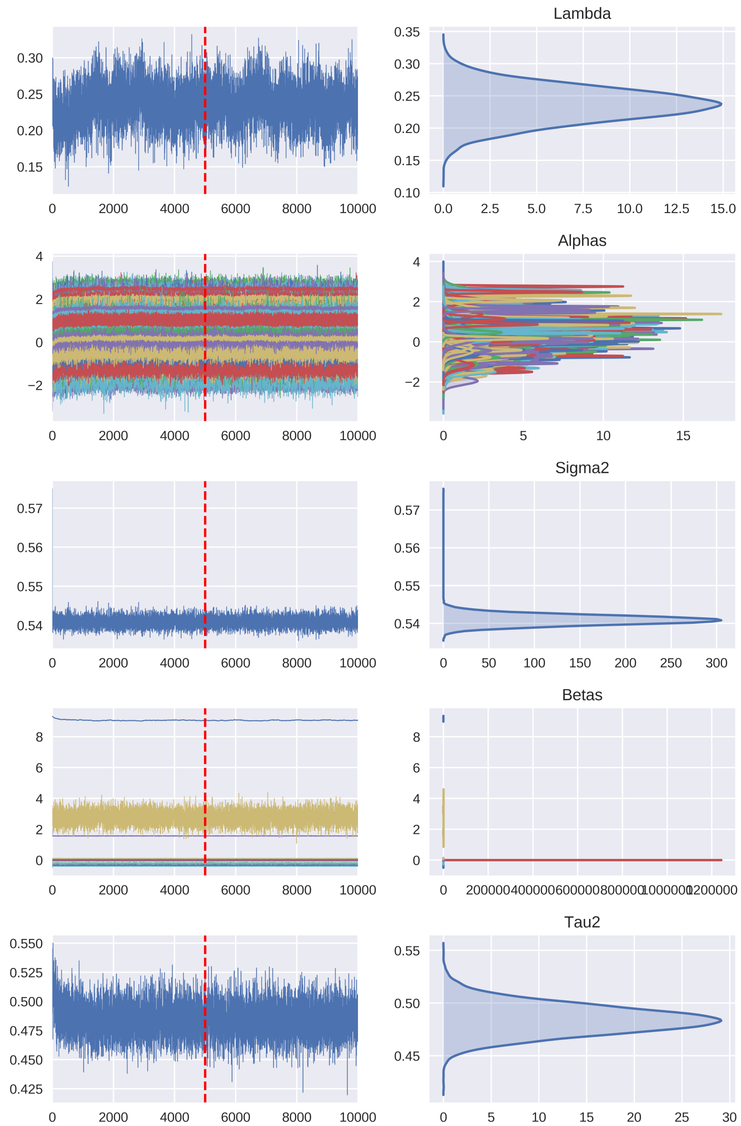


Figure 4.6: Trace plots (left) displaying simulated draws from MCMC chain for parameters and distribution of samples (right) for HSAR model. Vertical red line highlights the point along the chain where previous samples are discarded.

5 — Do visual-only features extracted from images of retail environments reflect different urban consumer experiences?

N.B. The research presented in this chapter is an adapted version of the publication: Comber, S. Arribas–Bel, D., Singleton, A. and Dolega, L. (2019) Using Convolutional Autoencoders to Extract Visual Features of Leisure and Retail Environments, *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2020.103887.

Abstract

Visual characteristics of leisure and retail environments provide sensory cues that can influence how consumers experience and behave within these spaces. In this paper, we provide a computational method that summarises the "visual features" of shopping districts by analysing a national database of geocoded store frontage images. While the traditional focus of social scientific research explores how drivers such as *proximity* and *attractiveness* of shopping environments factor into store patronisation and location choice decisions, the *visual* characteristics that describe the enclosing urban area are often neglected. This is despite a tacit assumption that consumers translate visual appearance of a retail area into a judgement of its functional utility, which mediates consumer behaviour, patronage decisions and the image a retail location projects to passers by. Such judgements allow consumers to draw fine distinctions when evaluating between competing destinations. We introduce a deep learning model known as Convolutional Autoencoders to extract visual features from storefront images of leisure and retail amenities, before partitioning these visual features into a sensible number of clusters. We then introduce measures describing the environment around the leisure and retail properties to differentiate between the clusters and assess which variables are distinctive for particular groupings. Our empirical strategy unpacks distinct groupings from the clusters, which implies the existence of relationships between visual features of shopping areas and functional characteristics of the surrounding urban environment. Overall, using retail environments, a core contribution of this paper seeks to demonstrate the utility of unsupervised deep learning methods to research questions in urban planning.

5.1 Introduction

Visual characteristics of urban spaces drive how individuals evaluate and experience their surroundings for the purpose of location choice behaviour and patronage decisions (Hauser and Koppelman, 1979). In the *The Image of the City*, Kevin Lynch argues the built environment can be drawn as "mental maps" that describe how the city is read visually by cues such as shapes, sizes and colours (Lynch, 1960). Not only this, Silver and Clark (2016) argue the actions, tastes, and traits of individuals create and support particular meanings attached to places. The measurement of a scene assesses the character of a particular place and highlights distinctive visual aspects of the built environment. As visual (but subjective) measures that describe scenes such as liveliness are hard to quantify with traditionally-available data, urban planners typically resort to building indicators that are based on more directly observable characteristics such as population density (Glaeser and Gottlieb, 2009) or street layout (Jung et al., 2017). Often representations that characterise the scenes of streets are inferred using visual audits conducted by researchers who collate data to explore similarities and differences of physical attributes visible from street-level – the quality of building facade, the presence of street art, or the condition of sidewalks, for example (Bader et al., 2017). Once aggregated, researchers can unpack relationships exploring the link between particular visual attributes of built environments and characteristics of the surrounding area. For retail environments, the visual image that shopping areas project to consumers is a function of a broad range of influences which affect patronage behaviour and consumer experiences (Bell, 1999). Retail area image is a multidimensional concept and to understand it requires unpacking the multitude of functional and visual characteristics that consumers associate with shopping areas (Baker et al., 1994). These characteristics are stimuli that influence consumer perception and, by extension, patronage intention for particular retail environments. Typically measures of retail area image are derived using survey approaches that rate characteristics such as the quality of building materials, the attractiveness of shop signage and overall environmental cleanliness (Bellizzi et al., 1983; El–Adly, 2007). As conducting in-person studies in shopping areas to record this data requires a high level of human judgement, they are cost-intensive and limited in the throughput required to construct visual descriptors of retail environments for large study areas

To circumvent the scalability issues of manually auditing a national sample of retail locations, we apply Convolutional Autoencoders (CAEs) to automatically extract visual features from images showing the frontage of leisure and retail properties across England and Wales. Particular interest on street-level imagery for leisure and retail amenities stems from their influence to the vibrancy of places and, hence, in the characteristics of the urban hierarchy (Dennis et al., 2002). While previous studies have shown that proximity to leisure and retail amenities factor into location choice decisions and patronage of retail environments (Glaeser and Gottlieb, 2009), the visual characteristics that describe the urban environment around the point of interest are neglected. Such approaches assume a “vacuum” around single amenities, which ignores the environmental context that surrounds these premises. As an example, the visual characteristics of a street with a restaurant accessible by several modes of transportation is likely to differ by the amount of liveliness when compared with another restaurant serviced in a location with no transport links. Capturing visual features of leisure and retail amenities allows an exploration into whether aspects of *what we see* are related to particular characteristics of the built environment that describe the amenities location. By clustering visual features extracted from the CAE, the principal contribution of this paper uses deep learning to assess whether visual-only features of retail landscapes correlate with observed characteristics of built environments, and whether there are distinctive characteristics for particular groupings.

The remainder of the paper is organised as follows. Section 5.2 motivates the underlying conceptual framework of the paper. Section 5.3 introduces the sources of data we utilise through the study, before describing the modelling approach we implement to arrive at our empirical objective. Section 5.4 presents the main findings. Finally, section 5.5 concludes the paper.

5.2 Background and Motivation

5.2.1 Visual characteristics of built environments

In the *Critique of Judgement* Immanuel Kant first observed aesthetic perception as a self-organising process that drives how individuals react to different environments (Kant, 1790). Not only do humans perceive their environment as neutral facts and data, but we react to distinctive aesthetic cues encoded in our surroundings that change how these spaces are experienced as we walk through them (Silver and Clark, 2016). Our judgement of the elements in our surroundings are rendered as a totality, independent of the constituent parts. When we stroll through a “hip neighbourhood”, the avant-garde feel, boutique stores, and DIY atmosphere are not perceived as independent objects. This is because they collectively recall a particular way of behaving that is adopted from the tastes and preferences derived from the environment the individual chooses to surround themselves with (Merleau-Ponty, 2004). Thus, an environmental psychology influences how preferences for certain environments are driven by a multitude of interwoven factors. Jane Jacobs recognised this as early as 1960, emphasising the role streets perform in setting the visual scene of cities. In a critique of modernist planning policy, Jacobs (1961) argued that unifying design elements of urban spaces is short-sighted, as the interplay of their “bits and pieces” are central to supporting the diverse excitement that street scenes offer.

Visual cues are seen as discriminative features that influence perceptions and evaluations of urban spaces, and even when considering socio-cultural biases in aesthetic judgement, have been shown to affect the psychological state of their inhabitants (Quercia et al., 2014). Kelling and Coles (1997)'s *Broken Windows Theory*, for example, suggests cues of environmental disorder in urban appearance such as abandoned cars, litter, and vandalism drive a perceived breakdown of social order which, in turn, induce more severe forms of criminal activity. Beyond disorder places deviate from conventional form by appearing, amongst other things, transgressive, glamorous, or informal (Silver and Clark, 2016). Thus, a suite of evaluative dimensions are considered when characterising the visual attributes of urban spaces, with different environments reflecting different visual representations of tastes and values. Not only this, Massey (1991) argues these particular spaces are not static, but have multiple identities that are forged by ever-changing social interactions occurring between people within them. All together, these considerations highlight the complexities of capturing a signal that reflects the visual qualities of street scenes.

5.2.2 Traditional approaches for describing retail environments

As aesthetic descriptions of urban environments such as glamorous, lively or conventional are difficult to measure directly, urban scientists typically fall back to constructing indicators of the qualities that describe spaces such as shopping areas (Silver and Clark, 2016). In-person visual audits strive to unpack how the functional, physical and social characteristics of retail environments correlate to affective outcomes such as store patronage and location choice decisions. Survey techniques have an extensive history in urban planning research, and borrow from psychometric measurement models to infer latent traits through an aggregation of single items visible across the audit (Bader et al., 2017). In UK planning discourse, for example, concepts such as vitality and viability have long underlined 'health checks' of town centre areas, reflecting arguments in Jacobs (1961) that thriving places maintain a diverse range of uses, attract significant numbers of people, and sustain a continuing ability to attract investment (Ravenscroft, 2000). Thus, vitality and viability is typically inferred by aggregating multiple items such as pedestrian counts, diversity of amenities, or boarded-up windows that are sampled at points across different retail locations. In the retail literature, several examples aggregate sets of measures to describe visual characteristics of shopping spaces. Bell (1999), for example, shows environmental stimuli such as appealing store colours, attractive shop signs and fashionable product ranges constitute a 'visual amenity' that inspires consumer willingness to patronise a shopping environment. Moreover, El–Adly (2007), finds attractiveness attributes of shopping malls such as luxury, comfort and convenience drive different patronage motives amongst different shopper segments in the UAE.

Survey-based approaches are often required to describe the visual properties of urban environments due to the absence of accessible and high coverage quantitative data (Salesses et al., 2013). Traditionally, studies are undertaken by relying on a mix of personal interviews, street-level observations of visual appearances, and annotated video recordings by experts (Quercia et al., 2014). This manual review of material is an arduous task however, and requires considerable collective effort to distinguish amongst the variety of visual cues encoded in the images.

5.2.3 Deep learning approaches for describing urban environments

To evaluate visual characteristics of particular places, Convolutional Neural Networks (CNNs) that are 'trained' with human-labelled images of street scenes are increasingly used to automate the classification of the scenes presented by built environments. This new body of literature has been punctuated by emerging access to new sources of data that have been released by commercial providers and photo-sharing websites in open formats (Arribas-Bel, 2014). Providers such as Google Street View (GSV) and Flickr have opened up access to street-level imagery for researchers through Application Programming Interfaces (APIs), which have, in turn, been used to construct modern crowdsourcing platforms for collecting millions of user perceptions about particular places. Large quantities of human-labelled, street-level imagery have been used for training computer vision techniques. Zhang et al. (2018), for example, use a deep learning based approach to predict perceptions of neighbourhoods in Bejing, China along six perceptual indicators of safe, lively, boring, wealthy, depressing, and beautiful, before investigating which visual elements correlate to a particular perception. The study used street-level images collated by MIT Media Lab as part of the "Place Pulse" program, which by fall 2018 had collected 1,566,218 pairwise comparisons between 110,988 street-level images from 56 cities worldwide (Dubey et al., 2016). This crowdsourced data was made publicly available by Salesses et al. (2013), who originally used it to understand the effect of the built environment's visual features on perceptions of safety, class and uniqueness in the cities of New York and Boston in the United States, and Linz and Salzburg in Austria. Additional studies that use labelled GSV images include Liu et al. (2016), who detect shifts in city identities and urban form for 26 cities from Europe, Asia, and North America. Lastly, Seresinhe et al. (2017) trained machine learning models on 217,000 crowdsourced images from the "Scenic-Or-Not" online game that rates outdoor, natural environments on an integer scale (1-10) of its *scenicness*, and explores questions that ask which types of greenspaces are perceived as beautiful.

Unfortunately, a drawback of these supervised methods are the large sample sizes required to train the network which are often unfulfilled in real-life applications. Moreover, these approaches typically utilise a large, non-expert workforce (voting on crowdsourcing platforms) to construct massive volumes of labelled image data. This creates several challenges. Principal amongst these is the balancing between maintaining a swift and economical annotation process while ensuring the collected labels are accurate (Sorokin and Forsyth, 2008). More importantly, the user's interaction with the labelling task may be influenced by socio-economic and demographic factors. As urban experiences are highly socially constructed, different groups might engage with the built environment in different ways, meaning visual characteristics are highly particular to various socio-economic or demographic groups (Quercia et al., 2014). These challenges exist because CNNs are supervised, meaning they require the network to be shown labelled instances of images for learning the nuances between particular predicted outputs. An alternative approach to extracting features from street-level imagery are *Convolutional Autoencoders* (CAEs). CAEs are *unsupervised* approaches meaning they provide a self-organised means for learning the relationships between elements in the data without being shown labelled inputs. CAEs are advantageous because they provide a less data-intensive alternative to CNNs that does not require the user to assemble large quantities of labelled data for training the network.

5.2.4 Application of computer vision methods to retail environments

While many studies that apply deep learning have focussed on urban environments, to our best knowledge, no application of deep learning to explore visual characteristics of retail environments currently exists in the literature. This is despite the high suitability of computer vision methods for characterising the variance in image attributes between different shopping areas. Consumers with little experience of a store or environment may use perceptual qualifications of image, in addition to prices, as a proxy for the quality of goods and service provision (Bell, 1999). Stimuli that influence consumer perceptions of shopping area image are functional qualities but also the aura of psychological attributes aroused by the environment. Functional characteristics include convenience and accessibility of store or retail area location, parking availability, the range of stores and products offered, and proximity to residential neighbourhoods and workplaces (Baker et al., 1994; Chebat et al., 2010). Psychological characteristics relate to the "visual amenity" experienced by consumers in shopping environments. For example, previous research links store patronage decisions to visual elements such as architecture, shop signage and exterior design (Baker et al., 1994), but also factors such as cleanliness and even colour of store premises (Bellizzi et al., 1983). Thus, quality inference for shopping areas is a function of multiple influences that affect consumer decision-making choices.

Given the wealth of research that has already linked image attributes of shopping areas to consumer patronage, the focus of the present study moves away from an exploration of footfall. Instead, our main research direction focuses on characterising the different *visual* representations of shopping environments by functional attributes that describe the area in which the premise is located. In synthesis of these two attributes, we unpack different representations of the *scene* that particular shopping environments project to passers by. The "*scene*" of an environment reflects both the visual characteristics and configuration of leisure, services, retail and cultural life, and data describing amenities such as leisure and retail premises are windows that allow researchers to unpack these configurations (Silver and Clark, 2016). An understanding of different *scenes* from leisure and retail environments is an important exercise because it unpacks patterns of urban human activity and function. This is useful information for retail planners and urban management schemes because it raises awareness of attributes and image among particular areas. Public or private sector agencies might utilise this to rationalise investment decisions that allocate spend to promotional activities and place marketing campaigns for building the profile of shopping environments (Page and Hardyman, 1996).

The visual design of retail environments are among the tools used to enrich the consumer shopping experience. Visual design of shopping areas has been manipulated previously to evoke desirable responses, such as arousal and pleasure which triggers approach behaviour and supports store positioning (Ballantine et al., 2010; Baker et al., 1994). Yet the visual design of retail environments in the UK is highly particular, and so consideration to its nuances is required for understanding potential implications to our applied methods. One limitation in applying computer vision methods to UK high street environments is a phenomena known as *clone towns* (Ryan-Collins et al., 2010). The idea of 'cloned' streets relates to the loss of identity and local character when chain stores come to homogenise high street environments at the expense of independent stores (Carmona, 2015). The implications for computer vision approaches concern the difficulty in identifying different typologies where no unique characteristics are directly observable from the images when they broadcast no local distinctiveness. Despite the British Retail Consortium (2019) arguing there have been calls for communities to reclaim their local high streets through the encouragement of local spending, it remains that a large number of distinctive facades constructed from local building materials may have been exchanged by identical glass, steel and concrete frontages (Ryan-Collins et al., 2010). This potentially limits the discovery of more interesting, diverse and distinctive types derived from empirical exercises that use street-level imagery from UK high streets. Despite this limitation, for wider study areas than would be permitted by in-person audits, computer vision approaches allow us to unpack how visual features of leisure and retail properties relate to functional characteristics of shopping environments, and consequently, how we can characterise the scenes these places offer.

5.3 Empirical Strategy

Our approach to explore differentiation between visual features of leisure and retail premises is three-staged. Firstly, we extract visual features from images of leisure and retail premises using a computer vision algorithm. Secondly, we partition visual features into a sensible number of clusters using a bottom-up classification strategy. And thirdly, to differentiate between the clusters, we introduce variables that describe characteristics derived from the point of interest around the properties.

5.3.1 Data

To implement the methodological approach we require two principal sources of data described below. Our first source of data are street-level imagery of 314,542 retail, service and leisure properties across England and Wales. These images display the front exterior of the property that face onto the adjacent street or open space. Exterior images were collected by a large pool of surveying teams equipped with hand-held cameras from the Local Data Company (LDC) in 2015. Sample images are displayed in Figure 5.1, and are categorised row-wise by several variables introduced in Table 5.1. As a pre-processing step, each JPEG image is resized from $800 \times 400 \times 3$ to a $224 \times 224 \times 3$ pixel image for compatibility with the applied neural network architectures, before normalizing the RGB values (0-255) to a 0-1 range. These resized, normalised digital images are the 3-dimensional inputs (width, height, and colour channel) to the convolutional neural networks we introduce in Section 5.3.2.

While this data offers new opportunities, there are limitations of using street-level imagery for visual audit purposes. Channels that affect perceptions of built environments such as sound and smell are absent from pictographic representations, and so cannot be directly evaluated from the image (Salesses et al., 2013). Similarly, small items less visible to the human eye that vary over short periods such as litter, drug paraphernalia, broken glass, or cracked sidewalks are difficult to measure given street-level imagery represent a single snapshot in time (Bader et al., 2017). More specifically, given the principle concern for the LDC surveying teams was to photograph facade features of the store premises, measures related to sidewalks such as number of parked cars or shrubbery might be partially occluded in the image, despite contributing to the overall ambiance of the urban area. Despite these limitations, the LDC images remain a valid source of data for our purposes. This is because they simulate a virtual walk down the street that replicates an eye-level experience, and the large number of LDC images provides granular, unprecedented coverage that would be impractical (and cost-intensive) to obtain otherwise.

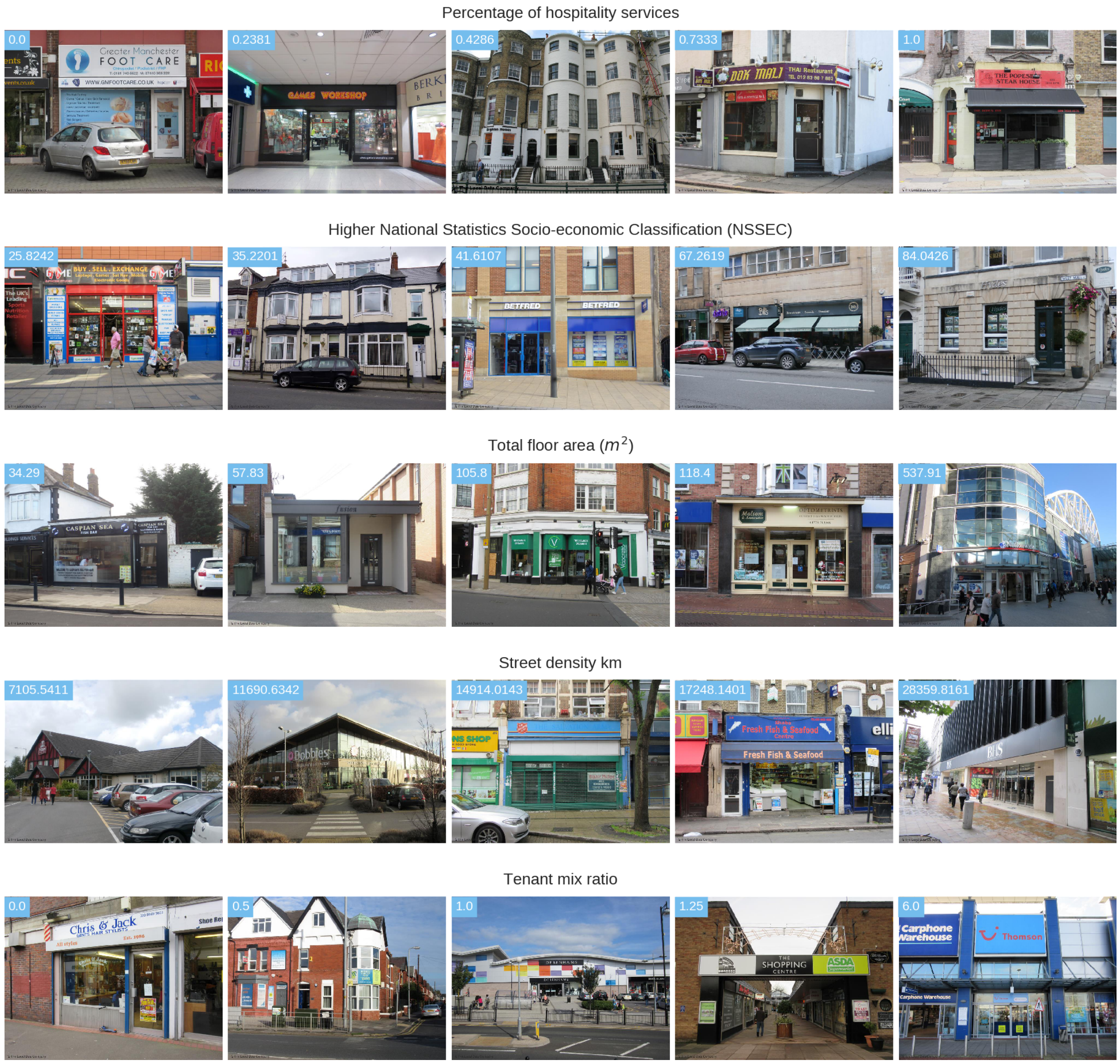


Figure 5.1: Sample LDC images for several features in Table 5.1. Each image is a random sample from each of five equal interval bins.

The second source of data is derived from characteristics that differentiate the particular visual representations of LDC images, and is used in the third stage of our approach. Our variable selection covers measures derived within a 15-minute walk catchment (assuming a walk speed of 4.5 kilometres per hour) around each leisure and retail premise (see Figure 5.2). These catchments are constructed using OSMnx, which is a Python library for acquiring, analysing and visualising street networks (Boeing, 2017). Within each catchment, we derive measures for a number domains outlined in Dolega et al. (2019) that describe shopping activity such as composition, diversity, size and function, and economic health (see Table 5.1). Aside from LDC and OSMnx data, we derive variables from several other sources. Census data is provided by the ONS (2016), our *e-res_score* variable is from a Consumer Data Research Centre (CDRC) data product and describes the vulnerability of town centres to the impacts of online shopping (estimated by Singleton et al. (2016)), and the *transport* variable is from the database of National Public Transport Access Nodes (NapTAN) (DfT, 2014). In addition, we use a small number of census-based socio-economic characteristics at Output Area (OA) level to describe the area in which the leisure or retail premise resides. OAs are built from postcode units and are the smallest statistical unit for which UK census data is published (ONS, 2019).

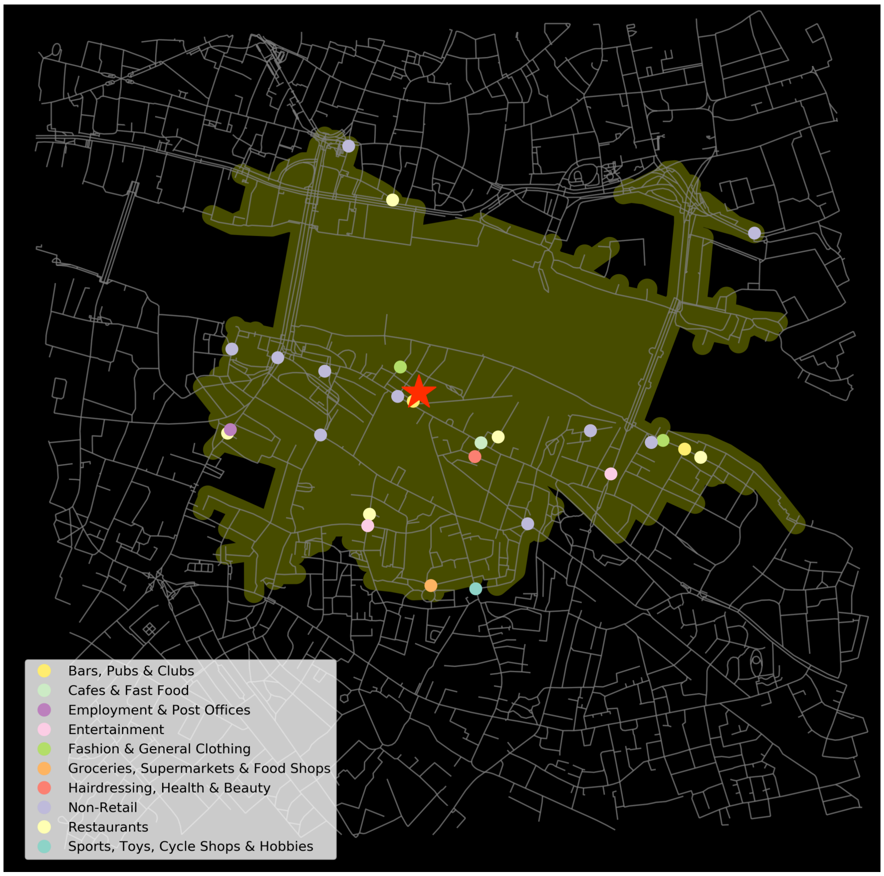


Figure 5.2: Example 15-minute walk catchment for a retail store around London Bridge. *Note*: 30 leisure or retail premises are sampled within the catchment to avoid clutter. Large red star denotes the store for which the catchment was created.

Variable	Description	Source	Mean	Std. Dev	Unit
Economic health					
bus_rate	Rateable value taxed on the business property.	LDC	33,113.86	146,627.15	Pounds
vac_rate	Vacancy rate of Local Authority District the property resides in.	LDC	0.09	0.03	Percent
unemployed	Percent of unemployed people in Output Area.	ONS	5.75	3.64	Percent
e-res_score	E-resilience score of nearest town centre.	CDRC	0.08	0.45	Score
transport	Number of bus or train links within catchment.	NaPTAN	61.21	38.19	Count
Composition					
comparison	Proportion of comparison goods stores within catchment (clothing, household goods, etc).	LDC	0.21	0.21	%
hospitality	Proportion of hospitality outlets within catchment (restaurants, bars, etc).	LDC	0.31	0.24	%
convenience	Proportion of food retailers within catchment (grocers, butchers etc).	LDC	0.13	0.18	%
consumer	Proportion of consumer services within catchment (banks, estate agents, etc).	LDC	0.18	0.21	%
tenant_mix	Retail to service ratio of catchment.	LDC	0.88	1.00	Ratio
store_diversity	Diversity of store types within catchment calculated by Shannon entropy.	LDC	1.14	0.57	Bit
Size and function					
floor_area	Total floor area for the property.	LDC	227.61	830.08	m^2
car_parking_spaces	Number of car parking spaces at the property.	LDC	1.28	17.36	Count
roeck_compactness	Compactness of catchment morphology.	OSMnx	0.49	0.13	Ratio
store_density	Number of stores within catchment.	LDC	14.24	20.45	Count
eig_centrality	Influence of store location within street network of catchment.	OSMnx	0.02	0.04	Score
street_length_avg	Average length of streets in catchment.	OSMnx	66.43	25.86	Meter
street_density	Total street length within catchment divided by catchment area.	OSMnx	15911.72	7134.42	Km^2
Socio-economic					
high_nssec	Percent of people with higher occupational employment in Output Area.	ONS	42.86	17.21	Percent
detached	Percent of housing units classified as detached in Output Area.	ONS	6.09	9.52	Percent
flats	Percent of housing units classified as flats in Output Area.	ONS	35.13	24.94	Percent

Table 5.1: Variable description for the domains of economic health, composition, size and function and socio-economics of leisure and retail premises.

5.3.2 Visual features from CAEs

Given the collection of leisure and retail property images are unlabelled and represented by a large number of raw pixels, a mathematical technique is required to decompose this larger set of correlated variables (or pixels) to a condensed set that captures the most salient characteristics of the image (Efron and Hastie, 2016). To learn this compressed set of variables from the raw pixels we rely on Convolutional Autoencoders (CAEs) (Goodfellow et al., 2016) which are composed of two layers: an encoder layer f_E and a decoder layer f_D. From a non-technical standpoint, the objective of CAEs is to take an input image, I, and reconstruct it as a copy, $\hat{I}$. Internally, CAEs use a hidden layer h that describes a code to reconstruct the image (Goodfellow et al., 2016). This lower dimensional mapping forces the CAE to prioritise aspects of the image that are the most useful for reconstructing a copy from the input image, meaning h learns the most useful properties of the data while discarding redundancies.

CAEs are extensions of autoencoders, which are techniques that essentially reduce the data under consideration to a smaller set of principal values. Practical applications of autoencoders include data compression for saving storage space and transmission times, and also cleaning corrupted data inputs by denoising. Thus, CAEs are autoencoders that introduce convolutional and (de)convolutional layers in the encoder f_E and decoder f_D sections, respectively:

$$f_E = \sigma(I * K + b) = h \tag{5.1}$$

where σ is a Rectified Linear Unit (Relu) activation function which is a truncation performed individually for every pixel x of the input, $Relu(x_{ij}) = max(0, x_{ij})$, that allows the CAE to learn non-linear patterns in the data, I are $224 \times 224 \times 3$ images where the 3 refers to the red, blue and green (RGB) colour channels, K are 3×3 matrices called convolutional filters, b is the bias unit which is similar to the intercept of a linear function and allows the line of the activation function to shift from the origin, and h is the code that represents the lower dimensional mapping of I. The convolution operator, $I * K$, is described more explicitly for the first layer in Eq. 5.2:

$$(I * K)_{xy} = \sum_{i=1}^{224} \sum_{j=1}^{224} K_{ij} \cdot I_{x+i-1,y+j-1} \tag{5.2}$$

which overlays each 3×3 filter over every possible pixel of the image, and records the sum of the element-wise product to an intermediate representation known as an activation map. The convolutional operator exploits spatial location in the image, as neighbouring pixels become activated for particular groups of edges that respond to semantically meaningful objects – trees, cars, or people, for example. This means particular filters become activated for specific patterns in the image, and stacking these filters across successive convolutional layers facilitates *parameter sharing*, where hierarchies of filters introduce levels of abstraction to the different kinds of features identified in the image (Goodfellow et al., 2016). As an example, the banks of filters learnt at the first convolutional layer might represent lower-level features such as lines, circles, and curves, while the higher-level convolutional layers will use these to construct whole objects – eye-like shapes or automobile wheels, for example. As the starting values of the K filters are randomly initialized, over the course of training the CAE the network will learn to find the optimal filter values that minimize the reconstruction error between I and $\hat{I}$.

Within each convolutional layer, a final step commonly applied to modify the output from Eq. 5.2 is pooling. In our case, after passing the intermediate representation through the Relu activation function, we apply the *max pooling* operation which returns the maximum pixel value within a 2×2 filter that steps across non-overlapping pixels of the input. This has the net effect of down-sampling an image by a factor of two, which sequentially reduces the pixel representation of our image from $224 \times 224 \times 3$ to a *latent* representation, h, which has shape $28 \times 28 \times 1$ and reflects the visual features we use for our clustering exercise (see Section 5.3.3).

To train the CAE end-to-end, we also require a decoder f_D network that reconstructs the original image $\hat{I}$ from h:

$$f_D = \sigma(h * U + b) = \hat{I}. \tag{5.3}$$

The only difference between f_E and f_D is that convolutional layers in the former are replaced by deconvolutional layers in the latter. This has the net effect of up-sampling the latent representation h ($28 \times 28 \times 1$) back to $224 \times 224 \times 3$, thus completing the reconstruction of the original image I. Once the CAE network has been sufficiently trained, the latent representation h, represented by $28 \times 28 \times 1 = 784$ pixels, becomes the basis of the visual features we use to differentiate between the visual scenes of different leisure and retail premises. To summarise these methodological steps, we visualise the resulting CAE architecture defined by Eq. 5.1 and Eq. 5.3 in Figure 5.3. In regards to implementation, the CAE model is defined in Keras (Chollet, 2015), with training undertaken on a single Nvidia Quadro M4000 GPU with 8GB memory. The Adam optimizer with default parameters is used to minimize the reconstruction error, which we evaluate using binary cross-entropy loss. Finally, we use a mini-batch size of 16 for 100 epochs, meaning updates to the convolutional filter's parameters are calculated using batches of 16 images at a time.

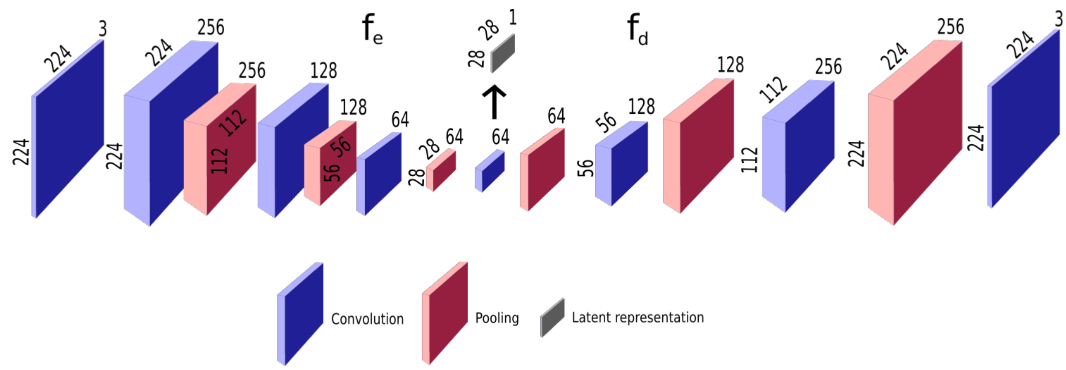


Figure 5.3: Convolutional Autoencoder (CAE) architecture showing encoder f_e, compressed representation h, decoder f_d and reconstructed LDC image $\hat{I}$. *Note*: filter numbers are shown horizontally along z-axis of feature maps, while width and height are shown along the x and y, respectively. Illustration was produced on the open-source vector graphics editor Inkscape (Inkscape Project, 2019)

5.3.3 Clustering visual features

To derive meaning from the visual features, we require a technique to group our vectors of visual features such that those in the same grouping exhibit similarities. This allows us to unpack similarities between the visual scenes for different retail environments which we can then describe by a number of functional characteristics outlined in Table 5.1. Our approach constructs a *bottom-up* classification where an initial typology with 250 numerous smaller groups are partitioned using k-means. Given the sensitivity of k-means to the initial starting values of the centroids, the algorithm is initialized 1,000 times with different centroid seeds, taking the final result as the output that best minimizes the within-cluster sum of squares. Finally, we allow up to 100,000 iterations within a single run to ensure stable convergence of the centroids. After the initial partition, we aggregate the clusters into coarser and larger groupings based Ward's method of hierarchical clustering (Ward, 1963). As Ward's method produces a dendrogram, we use it to slice a horizontal cut along the y-axis to create coarser levels of classification, which groups the 250 centroids of visual features into a smaller number of distinct clusters. This final partition represents the resulting clusters that differentiate the visual characteristics of the LDC images. Thus, we replicate a work flow similar to Spielman and Singleton (2015) and follow simple and widely supported methods to facilitate methodological transparency and reproducibility.

5.4 Results

In this section, we develop a discussion of our empirical findings based on two validation procedures. First, we undertake a validation exercise on our bottom-up clustering solution to ascertain a desirable number of clusters; and second, we explore consistency of group membership to particular clusters across sets of visual features generated from the CAE and two pre-trained CNNs. For brevity, the detailed outcome of these exercises are moved to Appendix 5.6.1 and Appendix 5.6.2. Based on the outcome of these exercises, in the following section we introduce several characteristics to unpack differences between the *five* distinct clusters of images we retrieve from our clustering approach.

5.4.1 Differentiating visual characteristics

To describe differences between the visual clusters, we aggregate characteristics for the consumer properties from Table 5.1, taking the median value for each variable per cluster[1]. To begin, we introduce radar plots in Figure 5.4 where each plot reflects a different visual cluster that shares similar psychological attributes reflected by common visual elements such as similar exterior design, signage, architecture, or colour. Along the axis of each plot aggregated variables that describe functional characteristics of these clusters are displayed. Thus, in synthesis of visual (psychological) attributes revealed by the cluster groupings and functional characteristics by the variables, we describe the *scene* projected by the clusters.

[1]Prior to the aggregation, we transform each variable to z-scores by standardization, $z = \frac{x-\mu}{\sigma}$, meaning each characteristic is rescaled by the fractional number of standard deviations from the mean value.

Figure 5.4: Median economic health, composition, size and function, and socio-economic characteristics in standardized units. Circular red line identifies zero, which shows standard deviations from the mean value.

Turning to the group sizes, we note the numbers of leisure and retail premises within the visual clusters vary substantially. Our largest cluster, Group A, contains 159,251 leisure and retail properties whose built environment is distinguished by high density street networks and large proportions of comparison retail outlets who sell merchandise that consumers purchase relatively infrequently and so evaluate prices, features and quality between stores before making a purchase. This includes outlets such as DIY & household goods, electrical, and clothing and footwear stores. Group A also contains a considerable proportion of hospitality outlets such as restaurants, bars and pubs, and entertainment venues. The Roeck compactness value measures irregularity in the shape of the retail area's boundary, with higher values indicating a highly compact retail area and lower values reflecting dispersion. The Roeck value for Group A, alongside its high street density, implies the urban morphology of the built environment around these stores is highly dense and not dispersed. All together, this suggests the scene characteristics of Group A reflects a bustling shopping area with relatively affluent residents who live in the immediate area (as shown by the high percentage of residents in higher occupational roles).

Group B contains 24,567 leisure and retail premises and is highly differentiated amongst its characteristics when compared to the other clusters. The functional attributes shared by leisure and retail premises inside this visual grouping reflect areas that have a low diversity of premise types, with the majority of outlets represented by comparison retail or consumer services such as car showrooms and house & home stores. Premises in this cluster are located in areas with high vacancy rates, meaning there are higher percentages of vacant or unoccupied store units relative to the other groupings. Moreover, outlets in this cluster appear to have high total floor areas and are serviced by fewer transport options, which conjures images of peri-urban spaces consisting of large retail units and warehouse spaces located on the fringes of dense urban areas and so are less beaming with consumer activity. Overall, the visual and functional characteristics of Group B portray a scene of sparse and less desirable retail and leisure land use when compared with the other clusters.

This is reinforced by socio-economic characteristics which reveal that individuals who live in the area, and might patron the shopping environment as consumers, typically occupy low percentages of high paid employment.

The next grouping that shares visual similarity is Group C, which contains 81,310 leisure and retail premises and is ascribed the label of 'Upmarket Hospitality'. The shopping environment of premises in this cluster are reflected by a large proportion of diverse hospitality outlets and leisure venues. This includes services ranging from restaurants and bars to theatres and galleries. A second defining characteristic of Group C is the extremely low vacancy rate when compared with the other clusters. This shows store units around the built environment for this grouping are typically occupied, which implies units in this cluster are in higher demand and so possibly elicit increased rates of rent. Similar to Group A, catchments around premises in this cluster are well served by transport links and possess highly similar urban morphology and socio-economic characteristics. In synthesis of visual similarities for leisure and retail premises within the cluster and functional characteristics of the urban landscape around these premises, Group C projects the scene of a thriving and upmarket shopping environment that is highly accessible and amenable to consumption activity.

Our smallest grouping, Group D, contains 6,962 leisure and retail units and is highly similar to Group C, although there are a few variables that differentiate the two clusters. Like Group C, Group D is characterised by a diverse range of hospitality outlets and stores that provide comparison goods such as electrical appliances and clothing. Compared to the dense street network of Group C, the urban morphology of Group D appears to reflect longer average street lengths that are fairly dispersed as shown by the low street density. Consistent with conventional wisdom, these two observations imply the built environment surrounding leisure and retail premises of Group D reflects high street shopping areas. Residents who occupy residential housing near stores in Group D typically occupy lower proportions of higher managerial roles. This suggests consumers, and by extension local consumption opportunities, are represented by less upmarket leisure and retail outlets given local patrons are typically less affluent than in Group C. Nine example images comparing low to high average street length for Group A and D, respectively, are shown by Figure 5.5. The presence of automobiles in images sampled from Group D suggest the built environment here is more amenable to vehicle use, with streets around leisure and retail premises in this cluster typically longer and less dense. All together, the composite visual and functional characteristics of Group D project a scene of long high streets that serve a diverse range of consumption purposes to local consumers.

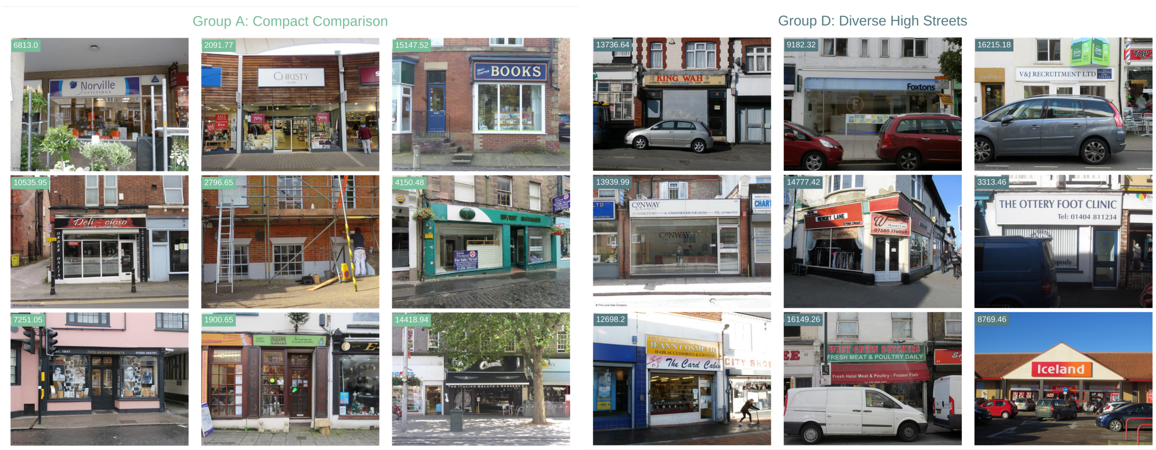


Figure 5.5: Leisure and retail storefront images and average street length values in metres sampled from Group A and Group D.

The last cluster, Group E, contains 81,310 leisure and retail premises and represents a middle ground between Group C and Group E. While units providing hospitality represent the highest proportion of services in this cluster, no particular mode of retail or leisure dominates unlike the other groupings. In fact, premises in Group E have the lowest proportion of comparison retailers in the surrounding urban environment. The urban morphology of Group E is fairly dense and compact, as evidenced by a relatively high street network density and Roeck compactness value. In synthesis, the shared functional attributes of premises in Group E suggest this grouping reflects a leisure, services and shopping environment that is accessed by consumers for everyday consumption as opposed to being accessed for a particular mode of retail or leisure service.

5.5 Discussion and conclusions

Visual characteristics of shopping environments are a significant determinant of area consideration and choice (Bell, 1999). Traditionally, visual representations of retail areas are retrieved using teams of human surveyors, who are cost-intensive to train and limited in the throughput necessary to construct the visual form of built environments. Consequently, in this paper, we use vast quantities of street-level imagery to explore whether visual features of leisure and retail environments correlate to measurable characteristics of built environments. This was achieved using a deep learning model known as Convolutional Autoencoders (CAEs) which learnt a compressed representation that captured the most salient characteristics required to reconstruct the image from a lower dimensional representation. Once these visual features were partitioned into a sensible number of clusters, functional characteristics that describe a 15-minute walk catchment from each premise were introduced to differentiate between the cluster partitions. By clustering the compressed representation, we were able to identify five partitions from the data that reflected different categorisations of the *scene* that particular shopping environments project to consumers across a national extent. This is important because information describing retail area image has historically been desired by retail planners for rationalising investment decisions in place marketing campaigns (Page and Hardyman, 1996), but is seldom available at wide geographical scales.

Furthermore, our findings unpacked patterns of retail activity and function, which demonstrated that certain visual features were distinctive for particular built environments.

From an urban planning perspective, the main implications of our study demonstrated that aspects of what humans *see* were related to particular functional characteristics of retail environments. This was a pertinent question for retail practitioners to ask, as while previous studies have shown that *proximity* to (and *attractiveness* of) amenities such as leisure plazas, galleries and shops enter into consumer patronage decisions (Glaeser and Gottlieb, 2009), the defining visual characteristics of these environments are typically ignored. This is despite visual amenity being an important influence on patronage behaviour and the *scene* that shopping environments project to consumers (Silver and Clark, 2016). A further contribution of the present study relates to several methodological innovations we introduce in the analysis. As our CAE model is unsupervised, it does not require large numbers of labelled images for training the model to produce visual features for each image. While the existing focus of the literature uses pre-trained or fine-tuned Convolutional Neural Networks (CNNs) for computer vision tasks in urban planning (Dubey et al., 2016; Seresinhe et al., 2017; Zhang et al., 2018), in the present paper we show that unsupervised techniques such as CAEs can also extract visual information from street-level imagery. This is advantageous for two reasons. Firstly, it does not require the user to assemble a large number of labelled images for training the CNN, which might possibly be derived from a non-expect workforce on a crowd-sourcing platform such as Amazon Mechanical Turk. And secondly, because pre-trained networks are often designed for a different purpose than that intended by the user, transfer learning approaches may provide sub-optimal performance if the images used are too heavily skewed compared to the data used to train the original network. Thus, while CNNs can be fine-tuned to the user's image data, a secondary contribution of this paper highlights the utility of CAEs for urban scientific tasks seeking to extract visual information from street-level imagery.

Despite these advantages, there exists conceptual and methodological limitations that frame the conditions for which the study should be interpreted. From a conceptual standpoint, it is reasonable to suggest the 15-minute walk catchment used to derive measures that describe the functional characteristics of the environment around each premise might not be reflective of reality on the ground. A 15-minute walk in a dense urban environment like London is likely to intersect a variety of scenes that possess polarised socio-economic and functional characteristics – for example, the short distance between the affluent and poorer areas of Clapham and Brixton, respectively. This means measures describing the built environment within each catchment might be inaccurate due boundary effects that influence area consideration and create barriers beyond which consumers do not patronize. From a methodological perspective, a further limitation is that repeatability of the empirical approach is conditional on the availability of suitable GPU hardware for training the CAE model end-to-end. Unfortunately, deep learning models require appropriate hardware to train, and this presents a financial barrier of access to researchers interested in replicating (or extending) the empirical strategy to their own datasets. Despite these concerns, the main contribution of this article presents directions for future researchers to employ the deep learning methods adopted by the paper. As CAE networks are unsupervised, they offer flexibility to researchers seeking to extract visual features from image data without using pre-trained networks. This is a pertinent point to consider because the target domains of pre-trained networks are often purposed to answer a different research question than that asked by the user.

5.6 Appendix

5.6.1 Appendix A: Cluster validation

The lack of a single global optimization procedure is an inherent limitation of clustering exercises, meaning the plausibility and usefulness of the classification are typically split between the purpose it serves but also a validation of its system-wide accuracy. With this in mind, we pair human intuition for ascertaining a sensible number of clusters alongside a metric used for measuring cluster compactness known as average silhouette width. To determine the quality of possible cuts to the dendrogram and, therefore, resulting number of final clusters, we calculate the average silhouette width for several partitions of the 250-class k-means solution. Silhouette width ranges from $-1 \leq s_i \leq 1$, with higher values being desirable as they imply low within-cluster dissimilarity; it is calculated as $s_i = \frac{b_i - a_i}{max(a_i, b_i)}$, where a_i is the average Euclidean distance of i to all other data points in the same cluster, and b_i is the Euclidean distance of i to the cluster nearest to the one i is assigned to.

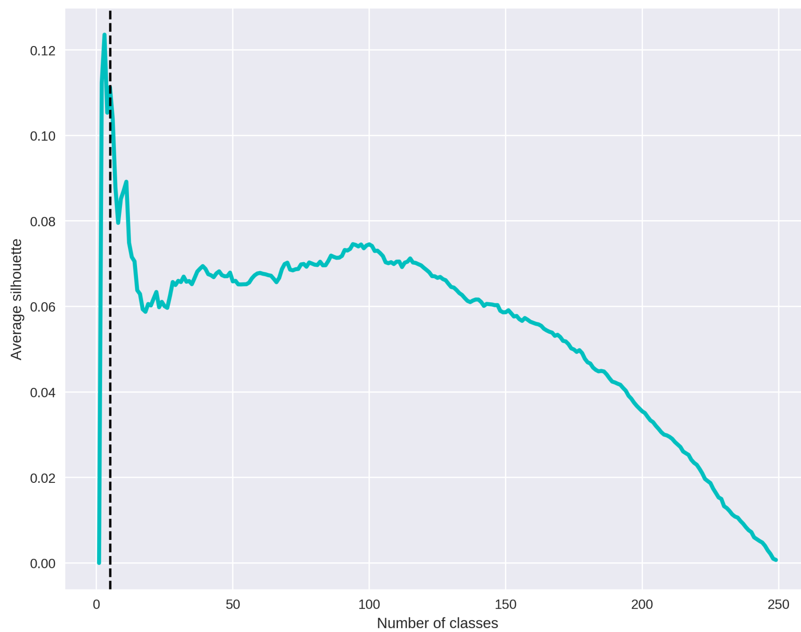


Figure 5.6: Average silhouette for different aggregations of the 250-class k-means solution. Vertical dashed line indicates the desired five-class solution.

In practice, we average s_i for all observations for each cut from 2 to 249 of the dendrogram in Figure 5.6, taking the final cut as one that yields a high average silhouette and sensible number of clusters. By scanning the figure we are able to discern a sensible number of five clusters which is ideal because five is both manageable to describe and large enough unpack interesting between-cluster variation. To accompany this, we provide the resulting dendrogram for the five clusters in Figure 5.7, which visualises the agglomerative steps used to aggregate the 250-class k-means solution into five coarser groupings. This is important because hierarchical clustering techniques do not provide cluster partitions automatically, and so *tree-cutting* procedures are required to return partitions that reflect similarities amongst observations in the agglomerative procedure. In our case, while other cuts to the dendrogram offered reasonable performance, we take the decision to cut the dendrogram horizontally at this particular position (of the y-axis in Figure 5.7) because the five cluster solution has a high average silhouette width and sensible number of clusters.

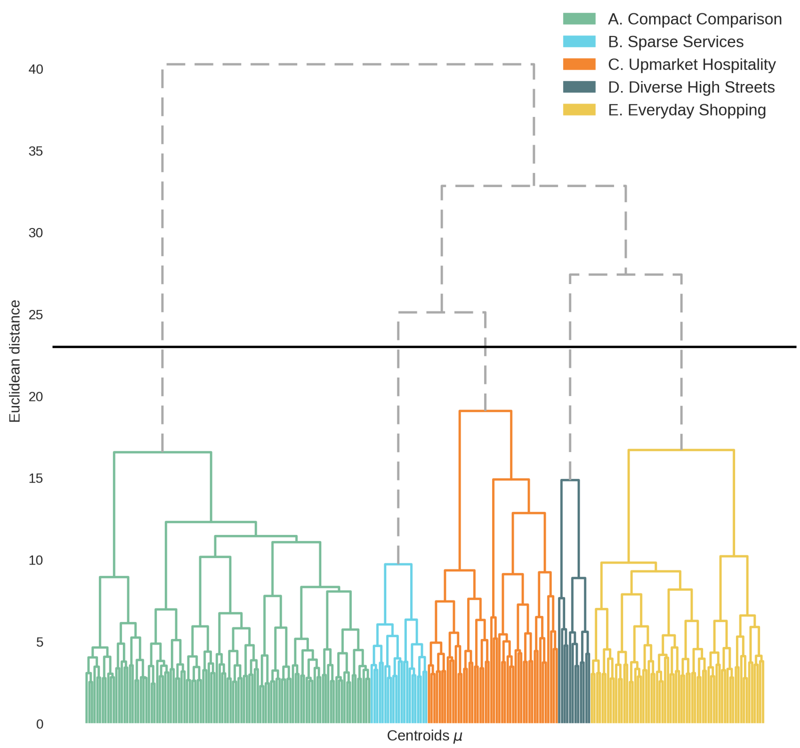


Figure 5.7: Dendrogram displaying the agglomerative merge of the 250-class k-means solution.

5.6.2 Appendix B: Consistency with pre-trained visual features

To benchmark the visual features, h, retrieved from the latent representation encoded by the CAE we extract a similar set of visual features from two pre-trained Convolutional Neural Networks (CNNs): VGG16-Places365 (Kalliatakis, 2017) and ResNet50 (He et al., 2015). While pre-trained CNNs are trained using large volumes of labelled data for predicting a pre-defined set of categories, CAEs learn visual information that is optimised to the dataset supplied by the researcher. Between these approaches reflects a trade-off between the generalisability of CNNs to extract features learnt from a larger pool of images and more focused visual information extracted from the CAE trained on the researcher's data. Irrespective of this, both serve as points of comparison to assess the consistency of group memberships to particular clusters across different sets of visual features. Given these networks are pre-trained, they are not required to be trained from scratch, and so are initialized with existing weights. For VGG16-Places365, the network weights are initialized to those trained on the Places365 database consisting of 365 different environment categories – highways, vineyards, or libraries, for example – and are tuned for scene recognition tasks. ResNet50, on the other hand, is initialized with weights trained on the ImageNet database, which is a large visual dataset consisting of hand-annotated images that represent a wider range of 20,000 categories. For these pre-trained networks, we remove the fully-connected layer at the top of the network, meaning instead of returning probabilities for categories, we extract the visual features that are discriminative towards particular categories instead. In all, three sets of visual features are introduced to the clustering exercise introduced below. This includes visual features from the CAE represented by 784 pixels, VGG-Places365 features by 512 pixels, and ResNet50 features by 2048 pixels.

To externally validate our empirical approach we monitor changes in group membership and cluster sizes between visual features extracted from our CAE and the two pre-trained convolutional neural networks (CNNs), VGG16-Places365 and ResNet50. Thus, after clustering each set of visual features from the three models, we explore *agreeability* of cluster membership for a five cluster solution in Figure 5.8. The cluster sizes are represented by the vertical white rectangles for the CAE, VGG16-Places365, and ResNet50 models (left to right), with the frequency of leisure and retail amenities changing between groupings shown by the stream fields, and so represent changes in the composition of clusters between the three models. From an initial reading of the figure a mixed picture emerges. While the group sizes are moderately consistent between the CAE and VGG16-Places365, the clusters formed from the visual features of ResNet50 are far more balanced, with leisure and retail amenities spread more equally amongst the partitions. In regards to group membership, the highest agreeability is observable between the largest clusters partitioned using visual features of the CAE and VGG16-Places365 models. Similarly, the clusters identified by '0' in both models seem to share moderate agreeability, with there also being minor agreeability between '2' and '4' of the CAE and VGG16-Places365 models, respectively; the frequency flows of the remaining clusters are far more dispersed between different clustering solutions. Agreeability with ResNet50 visual features, on the other hand, is observably low, with there being no discernible patterns and consistencies between the clustering solutions. This is unsurprising given the target domain of both pre-trained networks is highly dissimilar, a phenomena known as data bias (Chen et al., 2017). While VGG16-Places365 is optimized for scene recognition tasks, ResNet50 is trained to predict over 20,000 object categories from the ImageNet database, with classes ranging from particular types of plants to bedroom items. The weights of the ResNet50 network are tuned to generate visual features that are discriminative for a wider range of object classes, meaning when we recover a representation for each leisure or retail amenity image, the kinds of features activated are more generalised than those from VGG16-Places365. This is due to the narrow focus for the range of categories that VGG16-Places365 has been trained to identify (with an emphasis on scene recognition tasks), meaning the visual features are more likely to be similar to those derived from the CAE model. Therefore, as the

LDC images describe scenes observable from street-level, there is likely higher agreeability between the CAE and VGG16-Places365 models in terms of group membership and cluster sizes, which is reflected in the figure. All together, these observations confirm the visual features we extract using the CAE model are representing salient properties of the image, which motivates our descriptions for the characteristics of particular visual clusters in our empirical findings section.

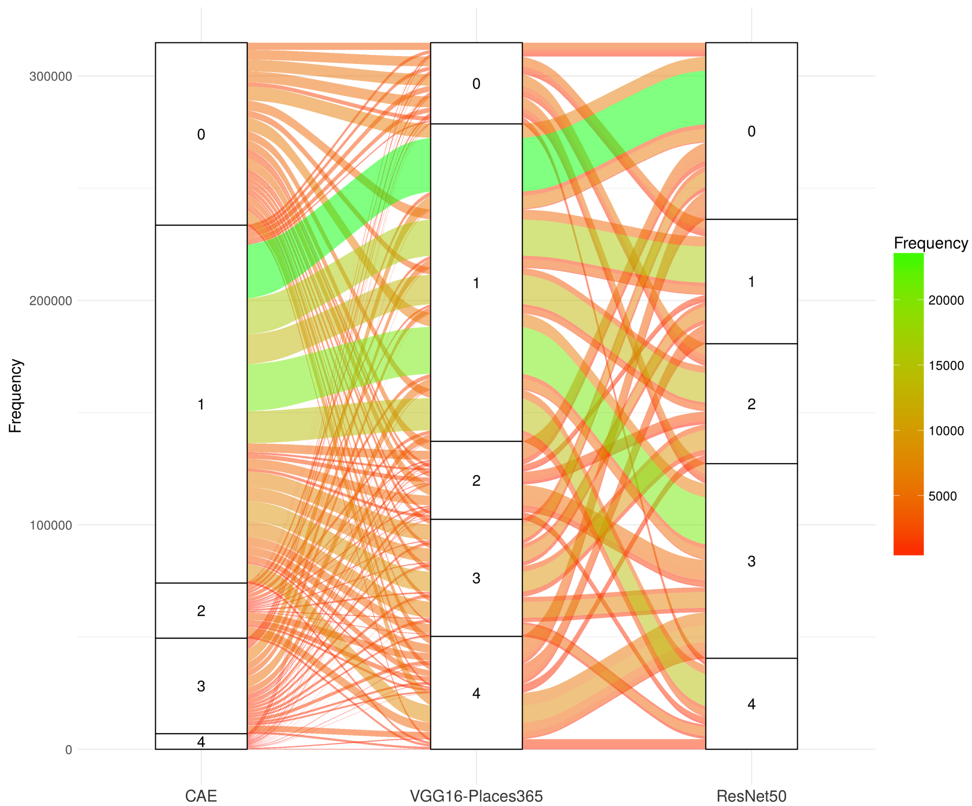


Figure 5.8: Agreeability of the five cluster solutions for visual features from Convolutional Autoencoder (CAE), VGG16-Places365, and ResNet50.

6 — Which physical characteristics of shopping environments drive the attractiveness of consumption spaces?

N.B. The research presented in this chapter was adapted to be submitted in the International Journal of Geographical Information Science, where it is currently under review.

Abstract

Quantifying relationships between physical characteristics of retail environments and consumer preference has long been of interest to urban planners and retail management. Attributes of environments such as shopping areas capitalise into home-buyer willingness-to-pay, and so reflect an important private benefit (or cost) for urban policymakers to consider alongside the wider spectrum of social and environmental factors. In this work, a deep learning approach is proposed to automatically detect the presence of various physical characteristics of shopping environments. A state-of-the-art computer vision model was used to detect instances of pedestrians, motor vehicles, pedestrian- and motor-orientated features among 249,190 street-level images of shopping, leisure and service premises inside a national sample of retail centres. By using these predictions within a regression modelling framework, we show how heterogeneity in physical characteristics across different retail centres influence preferences for particular shopping environments. Retail managers and place marketing initiatives might use this to establish commonalities between thriving consumption spaces, before employing findings to rationalise investment portfolios for public and private development. Ultimately our principal contribution demonstrates how object detection can be employed to automate the understanding of which physical environment properties drive the attractiveness of consumption spaces.

6.1 Introduction

Physical characteristics of street frontages are a central component of urban design because they contribute to the social life and vibrancy of public spaces (Gehl, 2010). Active street frontage is defined as the front exterior of buildings and includes physical characteristics such as windows, doors and presence of trees, but also pedestrian-orientated features like benches, parking bays and marked crosswalks (ODPM, 2004). According to best practice in urban design (Llewellyn-Davies., 2007), improving the condition of street frontages creates multiple social and economic benefits through increased property (and rental) valuations, enhanced civic pride, and improved pedestrian access (Heffernan et al., 2014). For shopping environments, the physical attractiveness of store frontages has been argued as an influential factor of consumer desires to patronise consumption spaces (Bell, 1999). Relationships between consumer behaviour and store image such as architecture and layout (Ward et al., 1992), attractiveness of shop signage (Dennis et al., 2010), and the blend of store colours (Babin et al., 2003) are all visual stimuli that influence the individual's experience while patronising a retail environment. Not only this, characteristics indicative of the attractiveness for shopping areas such as pedestrian facilities, tenant mixes, product ranges, traffic volumes and presence of vegetation and greenery are all drivers of preference for urban spaces (Borst et al., 2008; Teller and Elms, 2010).

Traditionally, measurement of perception towards urban features has been conducted using survey respondents for small numbers of sampled locations. Through stated choice experiments, researchers have previously utilized respondents to rate characteristics such as the prominence of people, vehicles and pedestrian spaces to evaluate perceptual qualities like the walkability, safety and vitality of urban places (Herzog, 1992; Borst et al., 2008). Yet, constructing a dedicated resource from in-person surveys that describe perceptual qualities of urban spaces is cost-intensive and limited in the throughput required to describe characteristics of large study areas. Often hours of in-class training for surveyors or volunteers is required prior to data collection, with studies often limited to particular districts of cities (Adkins et al., 2012). Most poignantly, sample designs lacking adequate coverage raise issues of external validity, which leads to speculation of how far findings generalize beyond the sample.

To circumvent these challenges, in the present paper we introduce a novel technique that has yet to be applied within the field of retail geography. More concretely, this involves a computer vision technology known as *object detection* to estimate revealed preferences for particular physical characteristics detected within shopping areas. Using an automated method for recording the physical characteristics of urban landscapes allows us to cover a national sample of 2,808 retail centres across England and Wales. Within these centres, we rely on street-level imagery consisting of 249,190 unique photographs for individual retail, leisure and service premises taken by Local Data Company (LDC) surveying teams in 2015. This allows us to extract characteristics such as the number of cars, people and benches from images of the store frontages for individual premises, before aggregating the detected objects upward to the retail centre-level. From this, we create a collection of features for evaluating associations between characteristics of shopping areas and their desirability from the consumer perspective. With this workflow in mind, we pose the following research question: which observable characteristics of consumption spaces increase (or decrease) the desirability of retail environments?

While significant advances in the use of computer vision techniques for recognising the image content of urban environments have been observed in the literature (Liu et al., 2017; Zhang et al., 2018; Ilic et al., 2019), there has been far less explicit attention to consumption spaces. We link a dependent variable that describes the attractiveness of shopping environments to its physical characteristics using retail centre willingness-to-pay (RWTP) values estimated by Comber et al. (2019). In doing so, this paper achieves three novel contributions. Firstly, our work is novel in exploring associations between consumer preference and physical characteristics of retail environments at the national scale. Secondly, our efforts introduce modern analytical tools that have yet to be applied to consumption spaces. A secondary intention of this paper is as a potential pedagogic tool to researchers interested in using computer vision tools for extracting visual information from urban environments. Finally, this work has important implications for urban policymakers. Understanding how attributes of built environments such as shopping areas influence value is critical, as this represents a private benefit (or cost) to be considered alongside the wider spectrum of social and environmental factors (Bitter and Krause, 2016). Empirical data describing observed physical attributes might be used by urban planners to establish commonalities between thriving retail spaces, with results used for directing public and private development, in addition to the selection from design alternatives. Overall, we find the presence of high footfall and pedestrian-orientated amenities were shown to be positively related to the RWTP for retail centres, while motor traffic and motor-orientated design features were found to be negatively related to attractiveness. Across a national sample of locations, our findings fill a research gap that use novel methods to address how physical characteristics of retail environments impact their desirability from the consumer perspective.

We organise the remainder of this paper as follows. In Section 6.2 we review related work and motivate the underlying conceptual framework of the paper. Section 6.3 introduces the empirical strategy used to detect semantic objects from the images and the approach used to determine associations between characteristics of retail centres and consumer preference. In Section 6.4 we introduce our sources of data, images displaying frontages of consumer space premises and variables at the retail centre-level that allow us to unpack relationships between objects detected and the willingness-to-pay for retail centres. Section 6.5 describes results of the proposed method before a discussion and concluding remarks in Section 6.6.

6.2 Background and related work

The public realm lays the terrain for social interaction, but also forms a significant part of the urban landscape's transaction base – high streets, market squares, shopping centres (Jalaladdini and Oktay, 2012), for example. Within retail environments, physical design elements that increase the attractiveness of a shopping environment have been argued to influence consumer location choice and patronage behaviour (Teller and Elms, 2010). Particular characteristics of shopping areas, such as the presence of leisure plazas or walkable spaces, may elicit positive consumer perceptions which increase the desirability of residential locations nearer to these particular retail centres. Matthews and Turnbull (2007), for example, demonstrate that proximity to consumption spaces that are pedestrian-orientated enhance local quality of life in the public domain, and so carry a positive effect on housing prices. This is because reinforcing pedestrianism strengthens the social function of urban areas, creating desirable consumer spaces that are sustainable, lively and safe. These arguments blend with the wider perspective in urban planning discourse that a multitude of valuable social, leisure and recreational opportunities emerge naturally when reinforcing life on foot (Gehl, 2010).

6.2.1 Perceptions of physical characteristics

Amongst the literature, different attributes of consumption spaces have been shown to drive the attractiveness of particular locations from the perspective of its users. Successful public spaces are often those with urban design characteristics that are comfortable, physically accessible and remove barriers to their use (Jalaladdini and Oktay, 2012). For example, pedestrian facilities inherent to lively streets such as bus stops and wide sidewalks have been shown to be positively related to the attractiveness of high streets (Borst et al., 2008). For older adults (aged 65 and over) in particular, street furniture such as benches to rest upon and pedestrian crossings by traffic lights have been shown to drive the attractiveness of walking, alongside increasing accessibility and perceptions of safety in shopping environments (Adkins et al., 2012). Alongside this, enhancing mobility choices through installation of bicycle facilities is another desirable design feature of lively, sustainable and healthy urban environments (Cervero et al., 2009; Gehl, 2010).

In more retail-focused terms, location desirability is an outcome of *spatial* and *non-spatial* characteristics that drive the attractiveness of shopping destinations (Rosiers et al., 2005). Briefly, non-spatial include factors such as: tenant mixes; merchandise ranges; staff friendliness; and value perceptions inferred from overall price, quality of products and price-quality ratios of merchandise, alongside advertised promotional offers (Baker et al., 2002; Teller and Reutterer, 2008; Teller and Elms, 2010). Spatial determinants include physical characteristics like site-related factors consumers evaluate when choosing between competing destinations. One of the most important factors includes accessibility, which comprises connectivity, signage and routing of the road network around the location, alongside parking conditions described by the availability and cost of spaces (Teller and Elms, 2012). Additional drivers of attractive retail environments include the physical presence of entertainment facilities like restaurants, bars, cinemas (Oppewal and Holyoake, 2004; Wrigley and Lambiri, 2014; Yavas and Babakus, 2009), in addition to non-shopping attractions like workplaces and transport facilities (Arentze et al., 2005; Teller and Reutterer, 2008).

High footfall is another characteristic that links to location attractiveness, in addition to reflecting a shopping destination's capacity to satisfy catchment needs and potential consumer spend (Mumford et al., 2020). Moreover, footfall is too a proxy of vacancy rates, which depends negatively on the number of passers-by. Shopping environments with a high level of vacancies typically induce lower volumes of pedestrian traffic and, by extension, decrease the already small number of non-vacant stores in the area (Koster et al., 2019).

Streets bustling with human activity, on the other hand, are typically considered attractive for walking, with busy streets often found to be wider and more amenable to pedestrian movement (Borst et al., 2008). Good walking opportunities are prerequisite for lively, safe, sustainable and healthy urban environments, with pedestrians deriving social and recreational value when reinforcing life on foot (Gehl, 2010). These factors relate to the vitality of urban environments which may influence shopper preference to particular retail spaces. As Jacobs (1961) famously remarked, higher pedestrian flows provide more "eyes-on-the-streets" which increases street-level sense of security, but also encourages more passive enjoyment of shopping spaces through people watching, social interaction and cultural exchange (Jalaladdini and Oktay, 2012). Pedestrian densities also tend to be higher in areas with a larger number of shops, with the majority of all pedestrian movement occurring through shopping (Koster et al., 2019). This stems from the positive shopping externalities derived from the clustering of retail, leisure and service premises. 'Trip-chaining' behaviour from consumers visiting several shops causes them to benefit by reductions in the transportation and search costs which are incurred when customers have to visit stores for each shopping trip (Claycombe, 1991). These positive shopping externalities derived from stores operating in close proximity are drivers of consumer preference (and utility maximisation) (Schulz and Stahl, 1996), which one might hypothesise to be speculatively internalised into valuations of nearby locations.

Lastly, several physical characteristics that detract from the atmospheric quality of consumption spaces have also been identified in the literature (Roggeveen et al., 2020). Exposure to noise derived from road traffic has been linked to increased likelihood of errors of perception and interference of communication (Barreiro et al., 2005) which, in retail environments, may degrade the visual and auditory atmospherics of the consumer shopping experience. Additional stimuli of unattractive retail spaces linked to motor vehicles are the reduced evaluations of safety from un-pedestrianised streets and air pollution emitted by motorised engines that detract from perceptions of environmental cleanliness (Teller and

Elms, 2010). Studies have shown retail units with easy access for vehicles have a negative influence on the attractiveness of walking, while a buffer between pedestrians and motor traffic is associated with increased perceived attractiveness (Adkins et al., 2012). Ultimately, these kind of people-place interactions have been documented across an impressive number of empirical studies (see Carmona (2019) for a comprehensive review).

6.2.2 Measuring perception in retail environments

Typically three ways of measuring the influence of physical attributes on consumer perception have been employed by practitioners: the traditional field survey-based approach, hedonic regression studies, and newer (automated) approach rooted in machine learning. Exploring associations between physical attributes and consumer perception has traditionally occurred through collection of data on-site, through physical observation or manually reviewing photographs taken at sampled points of interest. The observational unit of these approaches are qualitative judgements of particular characteristics ascertained from survey respondents, whose credentials might range from passers-by to expert assessors. In a relevant literature search, we found these studies typically use data points ranging between 288 to 2,139 respondents, with studies most often focusing on particular districts of cities (Juan, 2004; Borst et al., 2008; Teller and Reutterer, 2008; Yavas and Babakus, 2009; Teller and Elms, 2010; Adkins et al., 2012). Another line of enquiry evaluates perception from the perspective of revealed preference within a hedonic framework of valuation, where the implicit price of attributes are revealed by willingness to pay (Rosen, 1974). These methods unpack the value of complex goods as a function of its intrinsic and extrinsic characteristics. In related literature, retail premises can be viewed as complex goods within a hedonic modelling framework, where studies explore which attributes influence valuations of rental prices. A retail unit's market position can be inferred from its commercial rent (Hui et al., 2007); higher rents generally infer more attractive consumption spaces which draw consumers from wider geographical catchments due to the gravity of their composite retailers influence (Dennis et al., 2002). Alongside broader macro-economic conditions, exploring price determinants allow understandings of which physical attributes such as occupancy rates, gross floor area, and tenant mixes influence rental values and, mechanically, a retail locations brand/marketing position. Similar to before, a search of relevant literature found studies used between 151 to 4,738 premises, with most focused within particular neighbourhoods or city districts (Hardin and Wolverton, 2001; Mejia and Benjamin, 2002; Rosiers et al., 2005; Hui et al., 2007; Nase et al., 2015; Koster et al., 2019).

Small data studies encourage the necessity of theory and speculation of how findings drawn from samples generalize beyond the study (Lehmann, 2020). However, in this paper we argue leveraging modern analytical tools and big datasets provides a data-driven means to supplement and, where appropriate, create new knowledge that circumvents potential issues of external validity. This is because advances in digitization have allowed practitioners to conduct large-scale studies at micro level (Yin and Wang, 2016). Increasing access to digital media, whether as an *accidental* side effect of businesses moving online (Arribas-Bel, 2014) or industry-academic collaborations intended to create knowledge exchanges, has enabled researchers to access large sample areas while reducing time and logistical cost required for data collection. For instance, growth in the availability of street-level imagery, alongside adoption of deep learning algorithms, has enabled researchers to map visual information at fine spatio-temporal resolutions for wide geographical areas (Liu et al., 2017). The workhorse enabling these changes are Convolutional Neural Networks (CNNs), which are automatic image classification models that have been used for far ranging tasks. These involve the extraction of visual information from imagery, which includes tasks such as: pixel-wise semantic segmentation of urban scenes into different components – building, water features, sky, for example (Amirkolaee and Arefi, 2019; Helbich et al., 2019; Stubbings et al., 2019); quantifying neighbourhood characteristics such as liveliness, beauty and safety (Dubey et al., 2016; Naik et al., 2017; Zhang et al., 2018); and the automatic evaluation of quality and scenicness for building frontages (Liu et al., 2017; Law et al., 2018).

Despite the growing application of these methods in urban environments, often several photographs are required to reconstruct the scene of particular locations in full, with mobile elements such as pedestrians and motor vehicles potentially missed (Nasar, 1987). Alongside movement, channels such as sound, colour and smells present in urban settings are often absent from pictographic representations, and so cannot be directly evaluated from the image (Salesses et al., 2013). Despite these concerns, however, several studies demonstrate substantial correlation between responses to images and respondent opinions expressed on-site (Kelly et al., 2013). Moreover, as the principal focus of the present study is not of perceptual judgement, these concerns are far less grave. Instead, we are interested in identifying straightforward and objectively measured physical characteristics such as traffic lights and people, meaning street-level images remain a practical source of information for describing our consumption spaces.

6.3 Empirical strategy

The approach to explore drivers of consumer preference for retail centres is two-staged. Firstly, we describe the computer vision algorithm that is applied to generate features from the Local Data Company (LDC) images. And secondly, we introduce the modelling framework we use to estimate determinants of retail centre willingness-to-pay (RWTP).

6.3.1 Object detection network

To record physical characteristics from LDC images of consumer amenities inside our retail centres, we require an automated technique to identify objects relevant to our research objectives. We frame this problem as a computer vision task known as instance segmentation. This approach combines elements from classical computer vision problems of classification and semantic segmentation, but in addition, requires a correct detection of *individual* objects in an image. Thus, we differentiate between objects by precisely segmenting each instance into a fixed set of categories; this allows us to count object occurrences rather simply classifying whether (and where) objects exist in the image. To implement this, we run a forward pass for each LDC image through the pre-trained Mask R-CNN (Regional Convolutional Neural Network) (He et al., 2017) for detecting instances of relevant objects to our research agenda. Mask R-CNN is a state-of-the-art model designed by Facebook AI Research that is pre-trained on the Microsoft COCO (Common Objects in Context) database. Not only has it surpassed all prior state-of-the-art instance segmentation systems, but it conveniently classifies a number of classes relevant to our research problem, namely those relating to motor- and pedestrian-orientated features, footfall and motor vehicles[1] (see Figure 6.1). Passing our images through Mask R-CNN generates as many predictions as there are objects present, which we filter for relevancy and those that share a high probability of correctness. For brevity, we locate technical description of how we apply Mask R-CNN to our data in Appendix 6.7.1. As we highlight previously in Section 6.2.1, the object classes we extract from Mask R-CNN have previously been identified as stimuli of behaviour among consumption spaces and urban environments more generally. The presence of high volumes of footfall, for example, is implicit of good walking opportunities, with busy streets often more amenable to pedestrian movement (Borst et al., 2008). By automating the measurement of our four object variables, we uncover national-level reach that allows us to explore correlation between physical properties of consumer spaces and the perceived willingness-to-pay of these environments.

Through extraction of relevant objects, we construct variables that allow us to count

[1] The full list of COCO object categories relevant to our research are: cars, trucks, people, bicycles, motorcycles, buses, trains, traffic lights, stop signs, parking metres and benches.

the occurrence of detected objects and approximate conditions of consumption spaces. These variables are emboldened in Table 6.1, with several object classes combined into more meaningful, composite measures that reduce multicollinearity between features and capture latent factors that underlie built environments – pedestrian-oriented features such as benches and bicycles, for example. Counts of our four object themes – footfall, motor vehicles, motor- and pedestrian-orientated features – are then aggregated by taking their summation across each retail centre, and then dividing by the total area to return these four variables as densities per kilometre square.

Figure 6.1: Masks, bounding boxes and probabilities of objects detected from four sample LDC storefront images in Notting Hill, London.

6.3.2 Modelling framework

Our modelling framework relies on hedonic pricing models (Rosen, 1974), which treat "products" such as residential or commercial property as a differentiated bundle of attributes. In our case, we specify four different models that variably control for spatial and non-spatial price determinants of shopping environments, and use a hedonic regression approach to reveal the willingness-to-pay for particular characteristics identified by our object detection network. Our four models are specified at the consumption space scale as opposed to store-level because store clusters create positive shopping externalities through trip-chaining behaviour, which consumers seek via the bundling of wants and needs at a single location (Koster et al., 2019). Stores operate in a "co-opetitive", value-adding partnership, and the study of their benefits should be considered as a totality of the agglomeration rather than just a single store (Kotzab and Teller, 2003).

To estimate *revealed* preferences among our retail centres, our first model begins by specifying a baseline model that estimates the response as a linear combination of features (see Equation 6.1),

$$\log RWTP_i = \alpha + \boldsymbol{objects}_i \boldsymbol{\phi}_p + \boldsymbol{x}_i \boldsymbol{\beta}_k + \epsilon_i, \tag{6.1}$$

where $RWTP_i$ is the retail centre willingness-to-pay (RWTP) estimated value for each retail centre i in the sample, $\boldsymbol{objects}_i$ is a 1×4 vector of composite measures describing densities of objects detected for the ith retail centre per square kilometre, $\boldsymbol{\phi}_p$ is a 4×1 vector of coefficients that estimate the effect of detected objects on $RWTP$, $\boldsymbol{x}_i$ is a $1 \times k$ vector of control price determinants from Table 6.1, $\boldsymbol{\beta}_k$ is a $k \times 1$ vector of regression coefficients for controls to be estimated, and finally ϵ_i is the model residual term following a multivariate normal distribution $\mathcal{N}(0, \sigma_e^2)$.

6.3.3 Accommodating spatial autocorrelation

Our initial equation contains an assumption challenging our claim $\boldsymbol{\phi}_p$ is an unbiased and consistent estimator of $RWTP_i$. The inherent spatial nature of retail centres means a proportion of unexplained variance in $RWTP_i$ determination may relate to a spatial component (Anselin and Lozano-Gracia, 2008). To account for this, we improve our baseline model by modelling spatial effects of unobserved characteristics into a spatially autocorrelated error term, u_i, known as a spatial error model (SEM). The SEM is particularly useful in the presence of particular spatially-correlated omitted variables, such as store prestige. In this case, everything applies as in Equation 6.1, except the assumption of our error term ϵ_i being well-behaved is relaxed, as we now have the spatially-correlated error term u_i,

$$u_i = \lambda \sum_{j=1}^{n} w_{ij} u_j + \epsilon_i, \tag{6.2}$$

where λ is a scalar parameter ranging from -1 to 1 that captures the strength of spatial autocorrelation present in unobserved characteristics of retail centres, and w_{ij} is the ij-th cell of a spatial weights matrix, W. W is an $N \times N$ positive definite matrix that encodes the spatial arrangement of retail centres by assigning non-zero to pairs of retail centres that are neighbours, zero otherwise[2] A limitation of Equation 6.2 is that we impose a formal definition of spatial connectivity through W that might be incorrect. Our third equation specifies a spatial heteroscedasticity and autocorrelation consistent (SHAC) model (Kelejian and Prucha, 2007) that leaves the covariance unspecified, meaning we assume no

[2]Spatial connectivity at the retail centre level is subject to exponential distance decay expressed as,

$$W_{ij} = \begin{cases} 1, exp(-(d_{ij}^2)/d^2), & \text{if } d_{ij} \leq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

where d_{ij} is the Euclidean distance between retail centres and d is the fixed distance bandwidth of 26-kilometres which ensures that every retail centre has at least one neighbour.

functional form for distance decay. More formally, we apply a non-parametric estimator for the variance-covariance (VC) matrix that uses weighted averages for cross-products of residuals, with ranges determined by a kernel function (Anselin and Lozano-Gracia, 2008). Thus, individual rs-th elements of the non-diagonal spatial VC matrix $\hat{\Psi}$ are calculated as:

$$\hat{\Psi}_{r,s} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ir} x_{js} \hat{u}_i \hat{u}_j K(d_{ij}/d), \tag{6.4}$$

where subscripts index individual elements of the explanatory variables matrix X and residual vector $\hat{u}$, and K is a triangular kernel function that determines which ij-th pairs of retail centres are included in the cross-product calculation (Anselin and Lozano-Gracia, 2008). In our case, our kernel uses a variable bandwidth based on the distance to fifteen nearest neighbours. For further technical details on implementation of SHAC models see Kelejian and Prucha (2007).

6.3.4 Non-linear relationships

Finally, we compare our traditional (and linear) models with a modern approach originating from the machine learning literature, a random forest (RF), which allows us to explore potential non-linear relationships between RWTP and our four object variables (Breiman, 2001). This non-linearity can be visualised using accumulated local effects plots that show dependency between the response and individual variables by conditionally averaging over the values of all other features within “windows” around particular data instances. In our case, we grow 1,000 regression trees to maximal depth without pruning, also requiring a minimum of five samples within each leaf node. Fortunately, RFs require little manual tuning and are robust to constraints such as multicollinearity, making them suitable for our application. Once fit, graphical devices can be exploited to reveal insight into the underlying surface fit by the RF. Variable importance plots, for example, can summarize the relative importances of each variable, alongside the provision of a measure of variability in the predicted RWTP scores across the forest.

6.4 Data

Several data sources are required to implement our methodological approach. The principal source of data are 249,190 street-level images of every retail, service and leisure property in our sample that displays the frontages and physical characteristics of the premises' exterior. Images were collected by a large pool of surveying teams from the Local Data Company (LDC) in 2015, and were photographed using hand-held cameras across every major city and town in England and Wales. These 800 by 600 pixel images are accompanied by data describing features such as the store type, geo-location, number of parking spaces, and total floor area of the premise.

6.4.1 Retail centre-level variables

Our second data source are retail centre boundaries released as an open data product by Pavlis et al. (2018)[3]. Only stores located inside the boundaries of retail centres are used within the subsequent analysis. Examples of how stores nest hierarchically into retail centre boundaries are shown by Figure 6.2. For each retail centre we obtain an estimate of retail centre willingness-to-pay (RWTP), which represents the premium home-buyers attribute to proximity of a nearby retail centre. These RWTP scores are derived by Comber et al. (2019), and also reflect rankings of particular centres within a hierarchy, whose position relates to the size, attractiveness and gravity of their composite retailers influence. Centres

[3] https://data.cdrc.ac.uk/dataset/cdrc-2017-retail-centre-boundaries

that top the hierarchy offer multi-purpose, comparison shopping experiences that contain a wide variety of stores that cast a wide geographical reach on consumers. By contrast, low ranked centres are embedded in local economies and offer shops and services that are used by smaller consumer catchments. The attractiveness of retail centres to home-buyers are related to the composition and richness of shopping opportunity (Teller and Elms, 2012), and so centres that offer an amenity-rich environment elicit higher RWTP values (Comber et al., 2019).

Figure 6.2: LDC database provides a storefront image for every store (red circles) located inside the retail centre boundary. Example shown is for a retail centre in Birkenhead, Liverpool.

RWTP values for each retail centre allow an exploration of how different physical characteristics of shopping environments moderate the willingness-to-pay for retail centres. Thus, we compile a number of spatial and non-spatial characteristics that we hypothesise as determinants of RWTP. Broadly, these determinants align with four themes commonly found in the retail hedonic literature (Sirmans and Guidry, 1993; Nase et al., 2015): location-based factors, market characteristics, retail centre design, and customer drawing power. These themes cover several variables that describe characteristics of retail centres (see Table 6.1). Most variables are aggregated upward to the retail centre-level by taking the summation, average or density. Take, for example, the number of objects detected by our computer vision approach for images showing store frontages. By dividing the sum of detected objects for all stores inside the retail centre by its area, we create a collection of features that describe the density of physical characteristics per kilometre squared. Given the completeness of LDC retail premise records, this approach provides a realistic approximation of conditions that would be observed by in-person surveying teams.

While several variables described in Table 6.1 are intuitive, others deserve further clarification. Variables under the "Market and location-based factors" theme reflect dummy variables that describe classifications of retail centres derived from characteristics including composition, diversity, function and economic health from Dolega et al. (2019). Over forty variables that constitute these domains are used in their modelling process for building the classification. This complete set includes variables such as vacancy rate, store diversity, tenant mix and proportions of comparison, convenience, hospitality and consumer services, with the complete set found in the supplementary materials of their paper. As variation in RWTP caused by these attributes is accounted for in Dolega et al. (2019)'s classification, they are not re-entered into our modelling process. For *gvi_index*, we proxy street greenery by applying a colour recognition measure developed by Li et al. (2015) called the green view index (GVI) to individual images[4]. *average_circuity* reflects the average curvature of streets within a ten minute walking radius of the retail centre's centroid. Straight-line roads are more navigable, with evidence suggesting that a street morphology that is grid-like enhances internal connectivity and enables the viability of walking to satisfy a consumer's daily needs (Bitter and Krause, 2016). These catchments are also used as buffers for counting the number of transport options (rail, bus, and trams) within and

[4]We acknowledge this approach is problematic when separating street greenery (such as grass, trees and shrubs) from green objects like cars, green store decorations and even walls. This could be alleviated using image segmentation approaches, but given urban greenery is not our variable of interest, we leave this to future research.

nearby the boundaries of retail centres. All together, variables in Table 6.1 allow us to describe determinants of RWTP for 2,808 retail centres.

Variable	Description	Source	Unit
Dependent variable			
RWTP	Retail centre willingness-to-pay describing attractiveness of shopping environment.	Comber et al. (2019)	Natural Log.
Market and location-based factors			
local retail and service centres	1 for local retail and service centre, 0 otherwise.	Dolega et al. (2019)	Binary.
retail shopping leisure parks	1 for retail, shopping and leisure park, 0 otherwise.	Dolega et al. (2019)	Binary.
leading comparison leisure	1 for leading comparison and leisure destination, 0 otherwise.	Dolega et al. (2019)	Binary.
primary food secondary comparison	1 for primary food and secondary comparison destinations, 0 otherwise.	Dolega et al. (2019)	Binary.
traditional high streets	1 for traditional high streets and market towns, 0 otherwise.	Dolega et al. (2019)	Binary.
Retail centre design			
average circuity	Average curvature of street network within ten minute walk of retail centre.	OSMnx	Ratio.
gvi index	Average percent of street greenery in retail centre.	Own calculation	Percent.
pedestrian-orientated features	Number of marked crosswalks, traffic lights, benches and bicycles detected per kilometre squared.	LDC	Density.
car parking spaces	Average number of car parking spaces per premise inside retail centre.	LDC	Count.
motor-orientated features	Number of road traffic signs and parking meters detected per kilometre squared.	LDC	Density.
Customer drawing power			
motor vehicles	Number of cars, trucks and motorcycles detected per kilometre squared.	LDC	Density.
avg floor area	Average floor area of store premises inside retail centre (000s).	LDC	Count.
transport	Number of bus, tram and train stops within ten minute walk of retail centre.	DfT (2014)	Count.
footfall	Number of pedestrians detected in LDC photos inside retail centre per kilometre squared.	LDC	Density.

Table 6.1: Retail centre-level variables used in empirical approach. *Note*: items in bold represent variables of interest.

6.5 Empirical findings

In this section, we develop a discussion of the findings from our empirical approach in three directions: first, we explore frequencies and distributions of objects detected inside our sample of retail centres; second, we unpack findings of our modelling approach by describing which characteristics most significantly determine retail centre willingness-to-pay (RWTP); and third we introduce several graphical devices that describe the importance of particular variables for predicting RWTP and potential non-linear feature relationships between RWTP and our object themes.

6.5.1 Exploratory analysis of detected objects

Before exploring the frequencies and distributions of our four object classes, we first note how our applied computer vision technique allows us to extract valuable physical properties of consumption spaces that would be cost-intensive to attain otherwise. While our impressions are built from static snapshots of consumption spaces, we argue the scope and granularity of our data offers a national-level approximation of the inventories of footfall, motor vehicles, pedestrian- and motor-orientated features that would be impractical to obtain elsewhere. With this in mind, we first visualise the distribution of each object class for all retail centres in our sample (see Figure 6.3). Footfall appears to vary the most between retail centres, while the distribution of motor vehicles is far less spread, and has a slightly higher median number of vehicles when compared to the quantity of pedestrians detected. On the other hand, pedestrian- and motor-orientated features appear less numerable, with very narrow margins of variation across all retail centres in the sample.

Turning attention to particular retail centres, in Figure 6.4 we rank the top twenty by absolute number of footfall and visualise the proportions of object classes within each

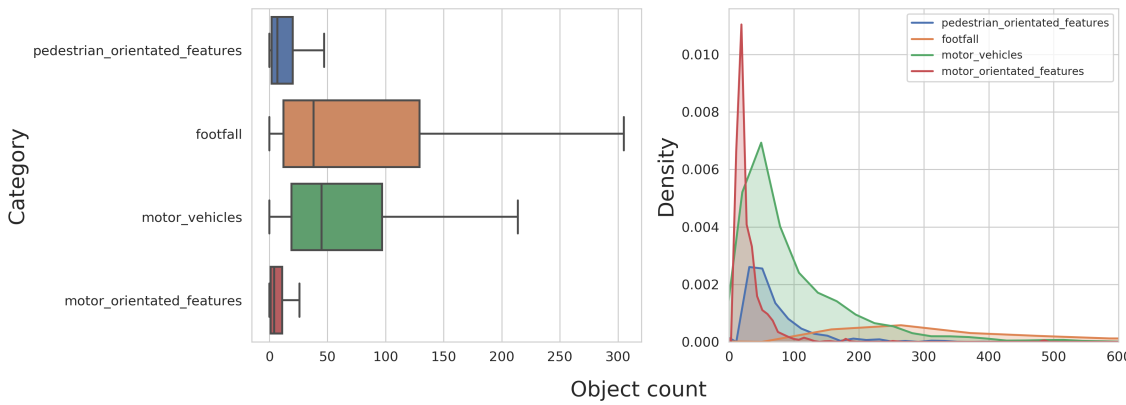


Figure 6.3: Distributional properties of footfall, motor vehicles, pedestrian-orientated features and motor-orientated features for retail centres in England and Wales ($n = 2808$).

centre. Unsurprisingly, retail centres in major UK cities such as London, Manchester, Newcastle and Liverpool occupy top rankings within the hierarchy, whose positions presumably relates to the size, attractiveness and gravity of their composite retailers influence. Interestingly, the proportions of each class are highly similar across the rankings, which indicates the likeness of human activity and urban design features that are detected from store frontages inside retail centres ranked highly by footfall. Of every retail centre in the ranking, The Hayes, Cardiff has the highest asymmetry of footfall compared to pedestrian-orientated features, which suggests consumers patronising this retail environment are serviced relatively less well by pedestrian facilities. On the other hand, Albion Street, School Close in Leeds appears to have the most balanced proportion of detected objects.

6.5.2 Determinants of retail centre willingness-to-pay (RWTP)

Interpretation of our specified linear models is performed by examining the magnitude, significance and signs of the estimated parameters, which we observe as generally consistent across each model. The main results for both the linear and spatial models are displayed in Table 6.2. First, we highlight reasonable goodness-of-fit, which implies the data could

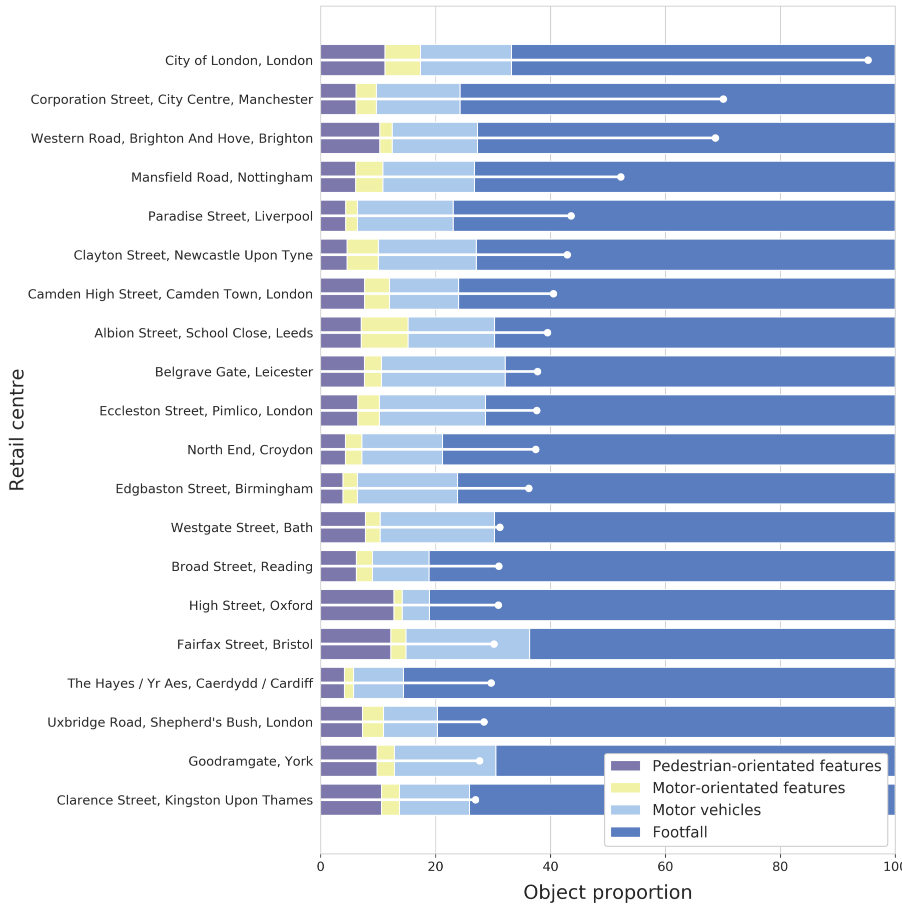


Figure 6.4: Percentage proportions of footfall, motor vehicles, pedestrian-orientated features and motor-orientated features for top twenty retail centres ranked by footfall (white horizontal line).

reasonably been generated under the fitted models. Each of the models have a similar Root Mean Square Error (RMSE) and pseudo-R^2, with the standard deviation of the residuals (or prediction error) ranging between 0.475 to 0.494 and around 49.8% of the variability in RWTP explained by our modelling approach[5]. We find a highly significant spatial parameter λ for the SEM, implying the presence of omitted variables that spatially correlate with the error term. While the standard and spatial models largely agree, the introduction of spatial effects to the SEM and SHAC in column (2) induces some marginal change in the magnitude of coefficient values. However, given the broad similarities between our models, we take the baseline OLS model specified in column (1) as our point of departure.

Before we begin, the magnitude of coefficients for variables in Table 6.2 represent semi-elasticities, and reflect the percentage change in the willingness-to-pay that consumers levy on residential property nearby the retail centre. To begin, we pay attention to market and location-based factors that are described by a series of categorical variables that reveal the RWTP for different typologies of retail centre. Compared to the reference category of 'leading, comparison and leisure destinations', retail centres classified as 'retail shopping and leisure parks' command the highest RWTP premiums of 88.4%. This suggests consumers are willing to pay 88.4% more than the reference category to reside closer to retail centres that have been identified as out-of-town locations which are typically occupied by 'big box' retailers and large multiple chains specialising in mass and value comparison merchandise (Dolega et al., 2019). By contrast, we estimate 'traditional high streets' as the least desirable consumption spaces of the retail centre typologies. This is presumably because these shopping environments focus more on convenience and local household services that are located in small market towns, traditional high streets or general rural areas, which is reflected in the 46.3% decrease in RWTP.

Turning to variables that reflect design characteristics of retail centres, we observe

[5] Potential multicollinearity of predictors were assessed using variance inflation factor (VIF) scores, and every variable was beneath the commonly-used threshold of five.

Table 6.2: Regression coefficients estimates of specified models. *Note:* variables in bold are rescaled for interpretability by multiplying coefficient by 100.

	Dependent variable:			
	RWTP			
	OLS	*SEM*	*SHAC*	*RF*
	(1)	(2)	(3)	(4)
(Intercept)	8.560***	8.703***	8.560***	-
	(0.286)	(0.282)	(0.321)	
Market and location-based factors				
local retail and service centres	-0.306***	-0.232***	-0.306***	-
	(0.033)	(0.032)	(0.045)	
retail shopping leisure parks	0.884***	0.923***	0.884***	-
	(0.052)	(0.050)	(0.076)	
primary food secondary comparison	-0.353***	-0.301***	-0.353***	-
	(0.033)	(0.032)	(0.035)	
traditional high streets	-0.463***	-0.449***	-0.463***	-
	(0.035)	(0.034)	(0.038)	
Retail centre design				
average circuity	0.374	0.258	0.374	-
	(0.261)	(0.256)	(0.292)	
gvi index	0.0081***	0.012***	0.0081**	-
	(0.025)	(0.026)	(0.036)	
car parking spaces	-0.031**	-0.038**	-0.031	-
	(0.015)	(0.014)	(0.024)	
motor orientated features	-0.069**	-0.071**	-0.069**	-
	(0.002)	(0.003)	(0.003)	
pedestrian orientated features	0.019***	0.027***	0.019**	-
	(0.007)	(0.007)	(0.008)	
Customer drawing power				
motor vehicles	-0.010***	-0.012***	-0.009***	-
	(0.002)	(0.002)	(0.002)	
avg floor area	0.035***	0.036***	0.035***	-
	(0.003)	(0.003)	(0.006)	
transport	-0.002***	-0.001*	-0.002**	-
	(0.001)	(0.001)	(0.001)	
footfall	0.024***	0.024***	0.024***	-
	(0.002)	(0.002)	(0.002)	
RMSE	0.475	0.494	0.475	0.313
pseudo-R^2	0.498	0.497	0.498	0.783

*p<0.1; **p<0.05; ***p<0.01

average circuity of the street network inside (and within a ten minute walk) of the centre as an insignificant determinant of RWTP. This implies consumer preferences are generally indifferent to whether the streets they shop amongst reflect a spatial ordering logic of circuitous and curving roads or more grid-like geometries. Regarding the extent of street-level urban greenery, we find Green View Index (GVI) as a significant determinant of RWTP, with a 0.81% premium for every one percent increase in GVI. This implies a *living* landscape element factors into consumer preferences for particular shopping environments. These findings appear to confirm previous studies that argue urban street greenery makes an important contribution to the attractiveness and walkability of built environments (Li et al., 2015).

Surprisingly, we estimate a significant negative relationship between RWTP and the average number of parking spaces at stores within the retail centre. While the convenience aspect of parking facilities have been identified previously as drivers of shopping destination attractiveness (Chebat et al., 2010), our data does not disambiguate by the type of parking facility or whether these spaces are free (Teller and Elms, 2010), which possibly accounts for the observed negative relationship. Next, we observe the density of motor-orientated features (such as parking meters, road signs, traffic lights and crossings per square kilometre) detected from LDC images inside retail centres as a significant determinant of RWTP. Every 100 additional motor-orientated features per square kilometre decreases the RWTP value of the retail centre by 6.9%. Conversely, increasing the number of pedestrian-orientated features per square kilometre by 100 increased RWTP by 1.9%. We lead a further discussion to qualify findings from these object themes in Section 6.6.

Finally, we interpret estimates of characteristics that describe our last theme, the customer drawing power of the retail centre. Consistent with conventional wisdom, the density of motor vehicles appears to have a negative relationship to RWTP. That is, for every additional 100 motor vehicles detected per square kilometre from the LDC images, the RWTP of the retail centre decreases by 1.0%. Regarding human activity, we find a significant positive relationship between footfall and retail centre attractiveness, with every additional 100 pedestrians detected per square kilometre inside the retail centres increasing RWTP by 2.4%. This finding emphasises the ability of attractive shopping environments to attract high volumes of consumer traffic. Next, we observe a significant positive relationship between the average floor space of retail, leisure and service premises and RWTP. These estimates reveal that consumers prefer store units that, on average, contain larger floor area dimensions able to stock a wider volume and array of merchandise. Somewhat counter-intuitively, the number of transport stops (bus, train and trams) has a negative, but small, association with RWTP. Every additional transport stop is estimated to decrease the RWTP of retail centres by 0.02%. Given transport facilities are expected to increase the consumer's ease of accessibility to the retail centre this appears surprising, but when we disaggregate the modes of transportation into single variables for buses, trains and trams a different picture emerges. While an additional train and tram stop increases RWTP by 2.7% and 7.9%, respectively, a single bus stops decreases RWTP, albeit marginally, by 0.03%. This reveals bus transit is a less desirable mode of transportation for consumers accessing retail centres.

6.5.3 Variable importance and non-linear feature relationships

Before embarking on a discussion of our object theme findings, we briefly explore additional insight derived from our random forest (RF) approach. Because the RF is a non-parametric technique it cannot be interpreted in similar terms as the linear models in Table 6.2. Instead, we use a variable importance plot and Accumulated Local Effects (ALE) plots to draw insight into the relevance and non-linearity of relationships between our explanatory characteristics and RWTP. Figure 6.5 visualises the importances computed from the RF in column (4) for every variable that enters our specification. The importance scores are ranked according to variables that yield splits in the decision trees among the forest that best reduce the sum of squared residuals criterion. The plot shows the majority of predictive power derives from just over two or three variables, despite the fact thirteen were used originally. While possessing a wide range of variability, average floor area dominates over the other variables, which implies stores where patrons are able to experience or shop amongst larger floor dimensions are valued highly by consumers. This is followed by whether or not the retail centre is classified as a retail, shopping or leisure park (*retail_park*), and the volume of footfall traffic generated.

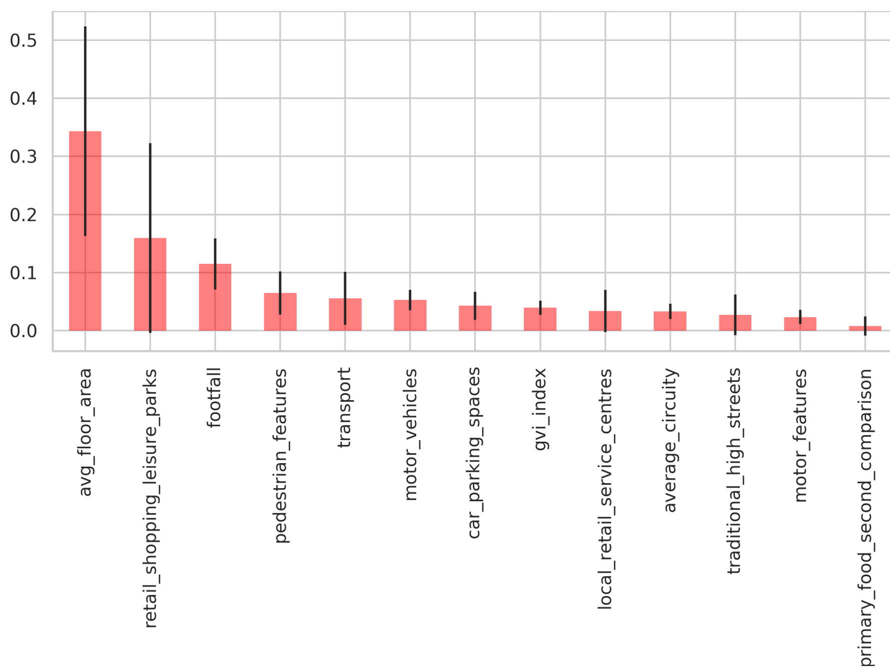


Figure 6.5: Importance of variables from Table 6.1 for predicting retail centre willingness-to-pay (RWTP). *Note*: black vertical lines represent variability of scores across trees in the forest.

To explore potential non-linear relations between our variables of interest and RWTP, Figure 6.6 displays the ALE plots derived from the RF for footfall, motor vehicles, motor- and pedestrian-orientated features. This figure highlights how the direction of relationship between objects we detect from our computer vision algorithm and RWTP changes across a reasonable range of values. This range of values is shown along the horizontal axis, where the effects of all other features are averaged within a small "window" of values, while the vertical axis presents values of the accumulated effect. Alongside this, we run a Monte-Carlo simulation of 100 replications where we randomly sample 60% of the data and re-compute the ALE, and we plot each replica as a light blue in Figure 6.6. Overall, there are substantial differences between each object theme. The first two variables in the top row display an overall positive relationship with RWTP, while the third and fourth variables in the bottom show demonstrate oscillation and some non-linearity, respectively. The accumulated effect function for pedestrian-orientated features shows an immediate marginal decrease, before remaining almost flat until 200 units, followed by a gradual positive increase across the remaining values. This evidence of non-linearity contrasts with the constant 1.9% increase in RWTP per 100 additional pedestrian-orientated features per square kilometre estimated from the linear models. On the other hand, the case of footfall is much more predictable, showing a constant linear increase. The relationship between motor-orientated features and RWTP is highly non-linear, oscillating majorly along the range of values on the horizontal axis. This behaviour contrasts wildly to the magnitude of coefficients estimated by the linear models, which imply up to 7.1% decreases in RWTP per 100 additional motor-orientated features per square kilometre. Finally, motor vehicles displays a negative relationship for lower values until around 800 vehicles per square kilometre, where the trend begins to saturate.

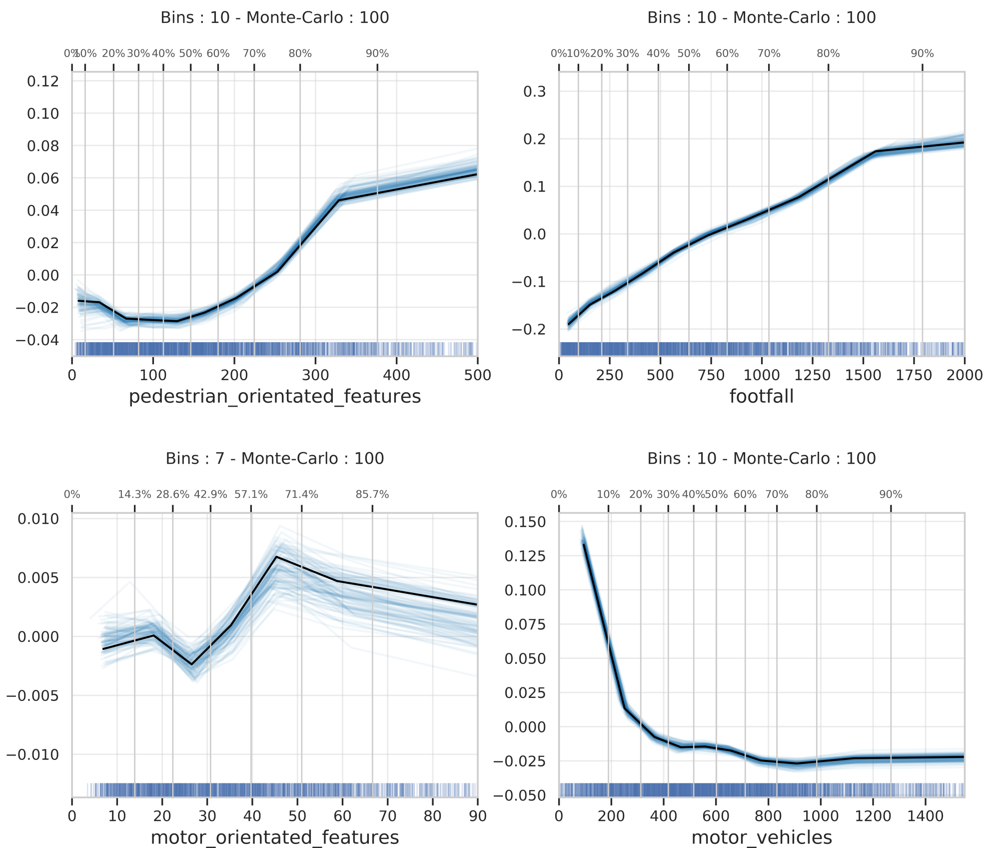


Figure 6.6: Accumulated local effects (ALE) plot for the four object themes. Monte-Carlo replicas shown by light blue lines. Rug plot of actual values that retail centres take for the variables shown by small vertical lines across horizontal axis.

6.6 Discussion, implications and concluding remarks

Under Glaeser et al. (2001)'s *consumer city*, the bundling of leisure, shopping and service amenities offered within consumption spaces induce positive externalities that are speculatively internalised into residential property located nearby attractive shopping spaces (Claycombe, 1991). Alongside this, physical attributes of consumption spaces themselves are also influential drivers of patronage intention (Nase et al., 2015), which our findings show to capitalise into *retail centre willingness-to-pay* (RWTP). Individual RWTP scores represent premiums attributed to the attractiveness of a particular consumption space (Comber et al., 2019), and using this measure with a simple linear model allowed exploration into which tangible characteristics of retail centres correlate to their perceived attractiveness.

For instance, one hundred additional pedestrians detected per square kilometre was found to increase RWTP by 2.4%, which aligned to previous work linking footfall density to the attractiveness, vitality and ability of consumption spaces to satisfy catchment needs (Koster et al., 2019; Mumford et al., 2020). Consistent with conventional expectation, a higher density of motor vehicles was found to decrease RWTP by 1.0%. Our result supports past evidence that shows automobile density as a driver of negative visual and auditory atmospherics within consumption spaces (Teller and Elms, 2010). Presumably, this accounts for the observed direction of relationship we discover, which has been recovered previously in hedonic studies that find negative relationships between home-buyer WTP and an increased presence of motor vehicles (Barreiro et al., 2005; Kim et al., 2010). Lastly, we found expected relationships between the density of pedestrian- and motor-orientated urban features. One hundred additional items of pedestrian furniture per square kilometre such as benches, bicycles (which are often detected as parked in bicycle facilities) and pedestrian crossings was shown to increase RWTP by 1.9%. This aligns to existing works that find urban design features that are comfortable, physically accessible and remove barriers to their use increase the attractiveness of public spaces, in addition to promoting more sociable and lively consumer environments (Gehl, 2010; Jalaladdini and Oktay, 2012; Adkins et al., 2012). Conversely, higher density of motor-orientated features was estimated to decrease RWTP by 6.9%. This might seem surprising, as features like parking meters are suggestive of convenient access. However, motor-orientated features are generally indicative of automobile dependency and high travel frequencies of cars, which likely detract from the overall visual and auditory atmospherics enjoyed by the consumer, in addition to the loss of social function (Teller and Elms, 2010).

In summary of our findings, this work has answered the research question that asks which observable characteristics of consumption spaces increase (or decrease) the desirability of retail environments? In doing so, we arrive at several novel contributions that are important within a range of contexts. Firstly, this study is the first to explore consumer preferences to physical characteristics of consumption spaces across a national picture of England and Wales. While previous works focus on perception within general urban landscapes (Dubey et al., 2016; Naik et al., 2017; Zhang et al., 2018), we pay particular attention to retail environments, and cover 249,190 retail, leisure and service premises located within 2,808 retail centres nationally. Second, we demonstrate a novel application of object detection within the fields of retail geography and urban economics. Our analysis automates the recording of shopping environment characteristics, and extends our coverage beyond which could feasibly be obtained from manual review of LDC photographs. Moreover, our efforts demonstrate how the complexity of deep learning machinery can be smoothed into a highly interpretable modelling framework. By linking object predictions to subjective preference, we show how deep learning methods can be used to communicate highly intuitive outcomes that are accessible to a range of audiences including policymakers, planners and retail managers.

While our findings are interesting themselves, more importantly they carry implications for decision-makers in urban and retail planning. Our findings provide information that strengthens the position of planners to justify the "public purpose" of investment portfolios. As we have shown, physical characteristics of retail spaces represent a private benefit (or cost) that, alongside the wider spectrum of social and environmental factors, are speculatively internalised into valuations of nearby locations. Equipped with these insights, urban and retail planners might use our findings to inform place marketing strategies, and allocate budgetary spend towards particular planning objectives that maximize successful consumer experiences (Page and Hardyman, 1996). More generally, our findings offer a natural succession to the seminal work of Glaeser et al. (2001). The *consumer city* phenomena emphasises how urban growth and development is coupled with consumption possibilities and urban amenities (Oner, 2017). The measure of RWTP for every retail centre allows us to approximate the attractiveness of these consumption spaces, but introduction of an object detection network enables us to go further. Recording physical characteristics within these spaces means we identify commonalities between attractive consumption spaces that drive higher valuations of nearby locations. This information is critical to policymakers, because it can inform how planners shape consumption spaces to maximize consumer experiences. If design considerations in retail spaces are optimized towards constructing the most attractive consumer environment, then local economies might see dividends from increased patronage which, mechanically, may incentivise higher spend.

Despite these potential implications, however, several limitations frame the conditions our study should be interpreted by. Foremost among these, our analysis presents aggregate drivers of consumption space attractiveness; this is despite consumption being highly individualistic. While our study yields unprecedented coverage, our aggregate approach likely smooths over more qualitative factors, such as the reasons why certain attributes render an environment to be more attractive. Such insight can only be revealed through qualitative approaches which, while covering smaller areas, are able to build knowledge from asking the finer questions. Moreover, the LDC photographs themselves represent a *literal* snapshot of consumption spaces, with several additional photographs required at each location to capture dynamism of the scene in full. While previous works show agreement between observational field and image-based audits (Kelly et al., 2013), the validity of our findings could be improved by collecting more photographs around the retail location, taking into account temporality of when photographs are taken, and otherwise moving beyond a non-causal empirical design. We leave this as an extension for future research.

6.7 Appendix

6.7.1 Appendix A: Mask R-CNN architecture

The Mask R-CNN architecture is conceptually simple, adopting a two-stage approach for classifying pixels into different object instances. The network's backbone consists of a Region Proposal Network (RPN) that contains two branches: a classification branch that outputs probabilities of 'objectness' and a regression branch that outputs coefficients describing coordinates of bounding boxes that contain objects. To begin, several candidate bounding boxes are proposed – these are known as *anchors*. As shown by Figure 6.7, these anchors are unevenly distributed across the pixel space and vary by size and aspect ratio. These candidate anchors are regions of interest that provide only rough localisations of objects, meaning refinement of these proposals are required to achieve precise localisation (Girshick, 2015). For each proposal, the RPN predicts a probability for an anchor being located in either the foreground (object) or background (non-object). Concurrently, the regression branch outputs four bounding box descriptors for each anchor that describe the x coordinate, y coordinate, width and height of a labelled box, $b = \{b_x, b_y, b_w, b_h\}$. As we use Mask R-CNN for inference, and because the neural network of the bounding box regressor is pre-trained, the parameters b_i for each anchor can be calculated from dimensions of the anchor (subscripted by a) and predicted bounding boxes outputted by the regression branch of the RPN (Ren et al., 2015),

$$b_x = (x - x_a)/w_a, \quad b_v = (y - y_a)/h_a, \qquad (6.5)$$
$$b_w = log(w/w_a), \quad b_h = log(h/h_a).$$

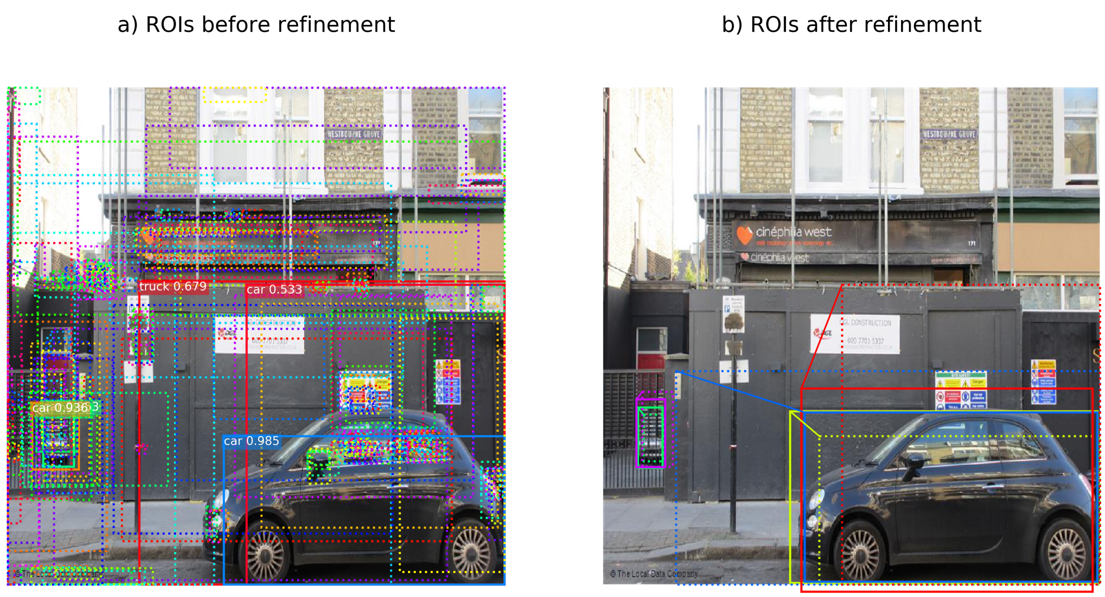


Figure 6.7: Demonstration of Region Proposal Network (RPN) refinement for LDC image. a) A random sample of 200 regions of interest (ROIs), or anchors, are displayed. Edges of proposal boxes classified as background and foreground are dashed and straight, respectively. b) Demonstrates precise localisation and removal of low IoU anchors.

Filtering for foreground objects only, the regressor descriptors b_i are then applied to the anchors for precise localisation. This moves the proposal bounding boxes to cover pixels of our images that are more likely to contain an object. A step known as non-maximum suppression is then applied on the remaining anchors. This step calculates the intersection of union for each anchor with every other anchor,

$$IoU_i = \frac{I_i \ \cap \ I_j}{I_i \ \cup \ I_j}, \tag{6.6}$$

where the highest-scoring IoU_i bounding box is retained from a group of overlapping boxes. A graphic that demonstrates these previous stages is shown by Figure 6.7. Finally, for each highest-scoring bounding box of objects in the image, an $m \times m$ segmentation mask that encodes the spatial layout of the object within the bounding box is generated on a pixel-to-pixel basis (see Figure 6.1 for example segmentation masks within highest-scoring bounding boxes) (He et al., 2017). Only detected objects above a probability threshold of 85% are included in our analysis, which we found the optimum threshold for maximizing the performance of our modelling approach.

7 — Conclusion

This thesis has introduced a data-driven framework to bring new answers and reinterpretations to long-standing theories common to the consumer spatial behavioural domain of retail geography. Through unprecedented access to data describing consumption environments, we illustrate novel means of empirically testing hypotheses exploring how consumer tastes and preference manifest in physical space. Among existing works, traditional efforts of testing research questions are often grounded in inferential methods designed to draw inference from small data samples, exhibiting well-behaved statistical properties (Kitchin, 2016). More recently, non-traditional sources of data emanating from the increased digitisation of retail environments have emerged from businesses expanding their economic model into digital product offerings. In some cases, this has facilitated the archival of large volumes of data reflecting many aspects of retail environments, and when this landscape is made available to researchers, it presents opportunities to study problems at degrees of detail and scope unthinkable only a few years ago (Arribas-Bel, 2014).

Taking advantage of this change, in this thesis we embraced an industry-academic partnership with a retail intelligence company who audit every consumer location in the UK, known as the Local Data Company (LDC), to access a nationally-comprehensive dataset describing retail spaces. This data describes a rich set of structured attributes relating to features such as type of consumer amenity and structural characteristics of the premise, but also unstructured attributes describing the text representations of the premise location and storefront images visualising the property's exterior. Completeness of the LDC's data afforded us a scale of spatial granularity at the national scale that is unseen within the present literature. This presented an opportune moment to revisit existing theories that explore the interface between retail environments and consumer spatial behaviour. Under existing works, theories and deductions of consumer tendencies are often predicated on coarse approximations of how actors behave, and further rely on small samples that were traditionally cost-intensive and limited in the throughput required to reconstruct the underlying empirical conditions of these complex systems. It is arguable that while our understandings of consumption spaces are premised on rich theories, their extent of truthfulness are contingent on study designs obtaining convincing out-of-sample generalisability and, mechanically, robust external validity. In traditional data-scarce contexts, these assumptions may not always hold.

Therefore, the central contribution of this thesis used the LDC dataset to unlock unprecedented access in retail spaces to empirically validate and, where appropriate, provide new insights to long-standing theories explaining consumer spatial behaviour. Yet, to decipher this picture our thesis also required the introduction of scalable methods to process new forms of unstructured data, such as storefront images and textual representations of business addresses. Often these sources are ill-purposed for traditional statistical analysis, and require powerful machine learning algorithms designed within the computer science community to extract useful signal embedded within the data. While innovations in machine learning, both unsupervised and supervised, have driven knowledge discovery in other fields of quantitative geography, retail geography has been slower to adopt these changes, despite the rich opportunity for potential cross-pollination. Therefore, a secondary contribution of this thesis served as a pedagogic tool to retail geography researchers interested in applying similar methods among their own works. In light of these two contributions, the following paragraphs provide an explicit answer to the four research questions proposed within this thesis, while also highlighting novelties of where non-traditional datasets and innovative methods have been used to reach these conclusions.

Our first research question asked: *to what extent can machine learning methods enrich* *data linkage for understanding retail environments?* To enrich data quality, record linkage is often a prerequisite step for identifying pairs of records that resolve to the same entity. More precisely, address matching is an enrichment exercise that allows practitioners to integrate disparate sources of data describing retail environments that would otherwise remain in isolation. In this empirical chapter, we demonstrated how machine learning could be applied to match two databases based on attributes encoded in the text representations of commercial addresses. More concretely, we resolved pairs of addresses from the LDC and Valuation Office Agency (VOA) databases of commercial premises using two recent developments in text-based machine learning – conditional random fields (CRFs) and word2vec – that had yet been applied for address matching. Using a supervised classification exercise, we found building comparison vectors for candidate pairs of addresses with CRFs address segmentation achieved a precision value of 95.5% and a recall of 90.2%. Meanwhile, our second approach augmented the first by replacing the string similarity metric used to compare segmented address fields with a comparison between word vectors. This augmented approach yielded a precision of 95.0% and recall of 87.0%. Interestingly, despite the increased sophistication of our word2vec-augmented approach, using hand-crafted features derived from domain knowledge of address structures outperformed the learnt features from word2vec. Regardless of this, the performance of both approaches contribute knowledge of how machine learning innovations in address matching show promising potential for facilitating the enrichment of data for use in downstream analytical tasks concerning retail environments.

Despite this contribution, however, we note several limitations that emerge from this empirical chapter. Firstly, the volume and scale of training data we obtained with a known match label is highly peculiar to this study. Accessing thousands of pre-matched addresses to train an address matching classifier is often an unrealistic expectation in many practical applications. Moreover, the Royal Mail's Postal Address File used to train the word2vec model is also a licensed product, whose access is contingent on a pricing subscription. Clearly, the outcomes of this study are contingent on data access that is often difficult to obtain, which highlights issues of reproducibility inherent to this research. While identifying these challenges, however, it is arguable these issues highlight gaps which future research in address matching might seek to fill. Recently proposed unsupervised and semi-supervised matching techniques have been successfully developed in the wider record linkage literature, and have reduced the manual effort of labelling data or eliminated it entirely (Jurek-Loughrey and Deepak, 2019). Extensions to this empirical chapter might seek to introduce these methods to address matching, and establish tooling that requires far less proprietary data to resolve pairs of addresses to a match.

The second research question we posed asked: *do urban hierarchies reflect spatial configurations of attractive consumption spaces and retail agglomerations?* Hierarchies of shopping spaces have long been examined at small geographical scales by academics (Dennis et al., 2002) and commercial organisations (CACI, 2018) alike. Yet, existing works lack fine spatial granularity at the retail centre scale, and this absence of quantitative evidence describing performance has precluded effective policy formulation and decision-making (Astbury and Thurstain-Goodwin, 2014). In this contribution, we constructed hierarchies of consumption spaces across the entirety of England and Wales. These hierarchies were formed by estimating retail centre willingness to pay (RWTP) scores, which reveal the premium home-buyers attribute to proximity of a given retail centre. Our findings unpack rankings of consumption spaces across a national network, where positions within the rankings relate to the size, attractiveness and gravity of composite retailers influence. Using a validation exercise, we further show associations between these rankings and characteristics associated with prospering and thriving locations. For example, we found as the vacancy rate increased by 1%, the RWTP of the consumption space decreases by 2.2%, which is consistent with expectation that large numbers of vacant units deteriorate the vibrancy of the streetscape, revealing implicit signs of decay. By using a dataset that describes unprecedented coverage of consumer spaces, our evidence provides a direct validation of Edward Glaeser's theory of *consumer cities* (Glaeser et al., 2001), as we find urban hierarchies *do* reflect patterns of attractive consumption spaces. This is a useful contribution because it highlights how a data-driven workflow can be leveraged as supporting evidence of much-used qualitative descriptions regarding how retail hierarchies are configured.

Again, despite the chapter's usefulness, we note several limitations of our approach. Firstly, an estimation of RWTP was only possible for 2,951 of 3,253 retail centres due to data availability. Ubiquitous national-level coverage is contingent on successful linkage between LDC and VOA properties, which unlocked the core attributes required for the modelling process – the business rate and geo-location to identify which retail centre a store nests within. Clearly, in around 9% of retail centres this linkage was not facilitated at all, which decreased the coverage our indicator claims to hold. Moreover, while this research presents a replicable and generalizable blueprint for constructing retail hierarchies, it is likely this data would be difficult for other researchers to acquire. This caveat is a reflection of the current state-of-the-art in terms of data access and reproducibility, which should motivate further efforts to democratise data accessibility. Extensions for future work might look to replace more restricted retail unit data with residential properties instead, which could also be used to extract willingness to pay estimates for retail environments. A further extension might also involve the creation of a longitudinal measure of RWTP estimates from future VOA data releases. This is enabled by the VOA continuing to reassess business rates of commercial property on a five-year revaluation cycle (VOA, 2014). Conditional on the ratings list being released as an open data product, our measure has updateability over time, which would allow exploration into the temporal dynamics of how this five-year window alters the hierarchies we observe.

Our third research question asked: *do visual-only features extracted from images of retail environments reflect different urban consumer experiences?* Long-standing theories show how visual features of consumption spaces provide sensory cues that influence consumer experiences and behaviours. Visual atmospherics such as colour, brightness, size and shapes within retail spaces are all conditions that affect levels of stimulation from consumers, which have been previously shown to influence purchase intention and preference for particular consumption spaces (Bellizzi and Hite, 1992; Ward et al., 1992; Bell, 1999). To approach this research question, our contribution trained a deep learning model, a convolutional autoencoder, on many thousands of storefront images displaying the front exterior of each unique premise. This computational approach allowed us to summarise "visual features" describing consumer spaces across a national database of images. After clustering these visual features, we unpacked five distinct groupings which we differentiated by introduction of several different measures including variables describing the economic health, composition, size and function, and socio-economic properties of the environment within a 15-minute walk catchment of each premise. Our exercise found distinct groupings from the clusters, which implied the existence of relationships between visual-only features of retail environments and different urban consumption experiences. One visual cluster we labelled as 'sparse services', for instance, appeared to be characterised by low diversity of premise types and high rates of vacancy, which were also serviced by very few transport options, conjuring images of sparse and less desirable retail and leisure land uses. Regarding practical implications, these findings are useful because they detect visual commonalities between particular urban consumer experiences, which might be used by practitioners to deconstruct the features of what makes certain visual environments more amenable to particular uses of consumption spaces.

As before, despite this contribution, we note several empirically-driven limitations of our findings that mirror open research problems that remain unresolved in the computer science literature. Firstly, the average silhouette width, which is a metric for the validation of consistency within clusters of data, was found to be relatively low. This implied our visual clusters of storefront images were poorly separated. k-means typically has difficulty in clustering non-spherical data of varying sizes and density. Problematically, our learnt embeddings from the convolutional autoencoder were highly non-linear, which meant a pair of non-linearly associated points may not be close in high-dimensional space. This creates difficulty in effectively defining a cluster "centre" (Wang et al., 2015), which is possibly reflected by the low average silhouette width in our study. In addition, while we cluster a condensed representation of the LDC images ($224 \times 224 \times 3$ or 150,528 dimensions) to a 784×1 embedding, we still maintain high dimensionality, which affects the convergence of k-means. This is because in high dimensional spaces the algorithm becomes less effective at distinguishing between observation points. Aside from these technical points, the low silhouette width might otherwise suggest the differences between visual clusters of storefront images are fuzzy, and not separable by simply minimising some Euclidean criteria such as the sum of squared error. Therefore, extensions to this work might seek to refine the clustering approach used here, as k-means is typically most suited for partitioning of hyperspherical clusters. Alternatives might involve adding additional pre-clustering steps to reduce the dimensionality of features further, or using soft clustering methods such as Gaussian Mixture Models to ascertain probabilities of cluster membership, and probabilistically filtering images based on which cluster a data point is most likely to form part of. These limitations are not restricted to the work presented here, however. Ultimately, problems working with high dimensional image data are unresolved in the literature and extend beyond this presented empirical chapter. For this reason, these limitations should not downplay the contributions stated in the preceding paragraph.

Lastly, our fourth research question asked: *which physical characteristics of shopping environments drive the attractiveness of consumption spaces?* Finding which physical characteristics of retail environments represent benefits (or costs) that internalise into location value is critical to arriving at attractiveness optimised design considerations (Page and Hardyman, 1996). Traditionally, previous works have explored ties between retail environments and their attractiveness through either *stated* or *revealed* preference approaches.

Stated preference works are often limited by small sample sizes resulting from challenges in obtaining survey respondents, while revealed preference approaches by historic confidentiality concerns related with accessing retail premise transactions (Rosiers et al., 2005). To circumvent these highlighted issues, this contribution used a scalable objection detection network to predict instances of footfall, motor vehicles, pedestrian- and motor-orientated features from 249,190 storefront images of consumer amenities across England and Wales. Despite using a complex deep learning model, our findings remained highly interpretable, as we linked predicted objects to subjective preference through an econometric modelling approach. Using the previously estimated RWTP values, we linked our proxy of attractiveness to various physical characteristics detected within different consumption spaces. Our evidence found the presence of high footfall and pedestrian-orientated amenities were positively related to the RWTP for consumption spaces, while motor traffic and motor-orientated features were found to be negatively related. By using a consistently measured and nationally-comprehensive sample, our contribution's findings can be interpreted as convincingly generalisable. Planners might use these findings as an impetus for devising place marketing strategies, or even rationalising the allocation of budgetary spend towards planning objectives that maximize successful consumer experiences.

To balance our noted contributions, we further outline limitations that frame the conditions of these findings. Firstly, our analysis presents an aggregate consumer behavioural picture for explaining drivers of consumption space attractiveness, despite consumption itself being highly individualistic. One limiting aspect of our study is that while we obtain unparalleled coverage, we likely smooth over aspects that can only be yielded from other, more qualitative-based approaches. As powerful and scalable as our object detection method is, its purpose is not tasked for answering the more finer questions, such as: "what is it about particular pedestrian features that render the environment to be more attractive?" Clearly, in its current iteration, our empirical approach would be unable to yield these finely grained insights. A second limitation relates to our coverage of retail environments being based on *literal* snapshots of individual consumer amenities. The lens of the LDC surveying teams camera's frame our window of opportunity to inspect the physical characteristics of these spaces, and its unquestionable that multiple additional photographs at each individual amenity will be required to reconstruct these locations in full. However, while noting the imperfections of these images, we argue this source is not imperfect enough to preclude the detection of interesting properties from consumption spaces. This is due to two factors. Firstly, the LDC images exhibit national-level coverage that would be difficult to acquire by other means. And secondly, while imperfect, the images retain enough signal to derive useful insight into the underlying physical characteristics of consumption spaces. Future extensions of this work might seek to increase coverage at existing locations by using Google Street View imagery to better approximate the physical conditions of these spaces.

In conclusion, through the accumulation of four empirical chapters, this thesis has sought to introduce a data-driven epistemological framework to retail geography research. By pairing modern analytical tools with unprecedented coverage of consumption environments, our efforts have brought new answers and reinterpretations to existing theories that explore consumer perceptions of retail environments. Yet, while noting these aims and expectations have *generally* been met, we further note several possible implications for the future of retail geography research in this direction. Firstly, caution is advised on using non-traditional datasets that are not originally purposed for research. A major flaw of datasets collected as by-products of business functions are that they suffer from varying degrees of quality and representability, which potentially compromises results and misleads conclusions. In our case, storefront images were certainly not purposed for building inventories describing the physical characteristics of consumer spaces, despite our use of the data in this way. Considerations regarding the extent to which the researcher's 'non-traditional' dataset approximates the reality under study should, therefore, emerge at the forefront of empirical considerations. Secondly, leveraging powerful machine learning algorithms requires some basic programming skills to access these sources, but also to manipulate, mould and pre-process data into a representation amenable to quantitative analysis (Arribas-Bel, 2014). While many computational models used within this thesis were pre-trained from other research groups, significant programming was still required to assemble a workflow, meaning these approaches are not always simply packaged into a single line of code. While many of these resources are pre-built into existing libraries, being able to navigate a landscape constructed by computer scientists brings advantages in being able to apply state-of-the-art methods to problem cases that have yet to trickle down to the wider community. With these points in mind, if retail geography researchers are to employ data-driven methods, then conquering these barriers of access is a first step in the right direction. Conditional on these changes, retail geography has a promising future in the age of big data.

Bibliography

Abdullah, S., Markandya, A., and Nunes, P. (2011). *Introduction to Economic Valuation Methods*, pages 143–187. Hackensack, NJ, USA: World Scientific.

Adkins, A., Dill, J., Luhr, G., and Neal, M. (2012). Unpacking walkability: Testing the influence of urban design features on perceptions of walking environment attractiveness. *Journal of Urban Design*, 17(4):499–510.

Alonso, W. (1964). *Location and land use: Toward a general theory of land rent.* Cambridge, MA, USA: Harvard University Press.

Amirkolaee, H. A. and Arefi, H. (2019). Height estimation from single aerial images using a deep convolutional encoder-decoder network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:50 – 66.

Anderson, J. (1971). Space-time budgets and activity studies in urban geography and planning. *Environment and Planning A: Economy and Space*, 3(4):353–368.

Angelides, M. C. (1997). Implementing the internet for business: A global marketing opportunity. *International Journal of Information Management*, 17(6):405 – 419.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., and Weaver, J. (2010). Google street view: Capturing the world at street level. *Computer*, 43:32–38.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):153–166.

Anselin, L. and Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science*, 92(1):3–17.

Anselin, L. and Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, 34(1):5–34.

Arentze, T. and Timmermans, H. (2001). Deriving performance indicators from models of multipurpose shopping behavior. *Journal of Retailing and Consumer Services*, 8(6):325 – 334.

Arentze, T. A., Oppewal, H., and Timmermans, H. J. P. (2005). A multipurpose shopping trip model to assess retail agglomeration effects. *Journal of Marketing Research*, 42(1):109–115.

Arnold, S. J., Oum, T. H., and Tigert, D. J. (1983). Determinant attributes in retail patronage: Seasonal, temporal, regional, and international comparisons. *Journal of Marketing Research*, 20(2):149–157.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49:45 – 53.

Arribas-Bel, D. and Reades, J. (2018). Geography and computers: Past, present, and future. *Geography Compass*, 12(10):e12403. e12403 GECO-1139.R1.

Asch, S. (1952). *Social Psychology*. Englewood Cliff, N.J., USA: Prentice-Hall.

Aspray, W., Royer, G., and Ocepek, M. G. (2013). *Anatomy of a Dot-Com Failure: The Case of Online Grocer Webvan*, pages 25–35. New York, USA: Springer International Publishing.

Astbury, G. and Thurstain-Goodwin, M. (2014). Measuring the impact of out-of-town retail development on town centre retail property in england and wales. *Applied Spatial Analysis and Policy*, 7(4):301–316.

Azevedo, C. D., Herriges, J. A., and Kling, C. L. (2003). Combining revealed and stated preferences: Consistency tests and their interpretations. *American Journal of Agricultural Economics*, 85(3):525–537.

Babakus, E., Bienstock, C. C., and Van Scotter, J. R. (2004). Linking perceived quality and customer satisfaction to store traffic and revenue growth*. *Decision Sciences*, 35(4):713–737.

Babin, B. J., Hardesty, D. M., and Suter, T. A. (2003). Color and shopping intentions: The intervening effect of price fairness and perceived affect. *Journal of Business Research*, 56(7):541 – 551.

Bader, M. D. M., Mooney, S. J., Bennett, B., and Rundle, A. G. (2017). The promise, practicalities, and perils of virtually auditing neighborhoods using google street view. *The Annals of the American Academy of Political and Social Science*, 669(1):18–40.

Badorf, F. and Hoberg, K. (2020). The impact of daily weather on retail sales: An empirical study in brick-and-mortar stores. *Journal of Retailing and Consumer Services*, 52:101921.

Badrinarayanan, V. and Becerra, E. P. (2019). Shoppers' attachment with retail stores: Antecedents and impact on patronage intentions. *Journal of Retailing and Consumer Services*, 50:371 – 378.

Baker, J., Grewal, D., and Parasuraman, A. P. (1994). The influence of store environment on quality inferences and store image. *J Acad Mark Sci*, 22:328–339.

Baker, J., Parasuraman, A. P., Grewal, D., and Voss, G. (2002). The influence of multiple store environment cues on perceived merchandise value and patronage intentions. *Journal of Marketing*, 66(2):120–141.

Baldovin, T., Zangrando, D., Casale, P., Ferrarese, F., Bertoncello, C., Buja, A., Marcolongo, A., and Baldo, V. (2015). Geocoding health data with geographic information systems: a pilot study in northeast italy for developing a standardized data-acquiring format. *Journal of preventive medicine and hygiene*, 56(2):88–94.

Ballantine, P., Jack, R., and Parsons, A. (2010). Atmospheric cues and their effect on the hedonic retail experience. *International Journal of Retail & Distribution Management*, 38:641–653.

Barreiro, J., SÃ¡nchez, M., and *, M. V.-G. (2005). How much are people willing to pay for silence? a contingent valuation study. *Applied Economics*, 37(11):1233–1246.

Barrentine, A. (2013). libpostal: international street address nlp [online; accessed 21 march 2018]. Available at: https://github.com/openvenues/libpostal.

Bateman, I., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., and Loomes, G. (2002). *Economic Valuation with Stated Preference Techniques*. Cheltenham, UK: Edward Elgar Publishing.

Batty, M. (2012). Smart cities, big data. *Environment and Planning B: Planning and Design*, 39(2):191–193.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3):274–279.

Batty, M. (2019). Urban analytics defined. *Environment and Planning B: Urban Analytics and City Science*, 46(3):403–405.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., and Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, 214(1):481–518.

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563.

Bearden, W. (1977). Determinant attributes of store patronage – downtown versus outlying shopping centers. *Journal of Retailing*, 53(2):15–22.

Bell, D. (2014). *Location Is (Still) Everything: The Surprising Influence of the Real World on How We Search, Shop, and Sell in the Virtual One*. Boston, MA, USA: Houghton Mifflin Harcout.

Bell, S. J. (1999). Image and consumer attraction to intraurban retail areas: An environmental psychology approach. *Journal of Retailing and Consumer Services*, 6(2):67 – 78.

Bellenger, D., Robertson, D., and Greenberg, B. (1977). Shopping center patronage motives. *Journal of Retailing*, 53(2):29–38.

Bellizzi, J., Crowley, A., and Hasty, R. (1983). The effects of color in store design. *Journal of Retailing*, 59:21–45.

Bellizzi, J. A. and Hite, R. E. (1992). Environmental color, consumer feelings, and purchase likelihood. *Psychology & Marketing*, 9(5):347–363.

Benjamin, J. D., Boyle, G. W., and Sirmans, C. F. (1990). Retail Leasing: The Determinants of Shopping Center Rents. *Real Estate Economics*, 18(3):302–312.

Berry, B. (1963). *Commercial structure and commercial blight: Retail patterns and processes in the city of Chicago*. Chicago, IL, USA: University of Chicago Press.

Berry, B. J. and Garrison, W. L. (1958). Recent developments of central place theory. *Papers in Regional Science*, 4(1):107–120.

Beule, M. D., den Poel, D. V., and de Weghe, N. V. (2014). An extended huff-model for robustly benchmarking and predicting retail network performance. *Applied Geography*, 46:80 – 89.

Birkin, M., Clarke, G., and Clarke, M. (2002). *Retail Geography and Intelligent Network Planning*. Chicester, UK: Wiley.

BIS (2011). *Understanding High Street Performance*. Department for Business Innovation and Skills [Online; accessed 8 July 2020]. Available at: http://urbanpollinators.co.uk/wp-content/plugins/downloads-manager/upload/understanding-high-street-performance.pdf.

Biswas, D. and Szocs, C. (2019). The smell of healthy choices: Cross-modal sensory compensation effects of ambient scent on food purchases. *Journal of Marketing Research*, 56(1):123–141.

Bitter, C. and Krause, A. (2016). The influence of urban design packages on home values. *International Journal of Housing Markets and Analysis*, 10(2):184–203.

Black, J., Hashimzade, N., and Myles, G. (2017). *A Dictionary of Economics*. Oxford, UK: Oxford University Press.

Blanchette, C., Dekoven, M., De, A., and Roberts, M. (2013). Probabilistic data linkage: A case study of comparative effectiveness in copd. *Drugs in context*, 2013:212258.

Blecic, I., Cecchini, A., and Trunfio, G. A. (2018). Towards automatic assessment of perceived walkability. In Gervasi, O., Murgante, B., Misra, S., Stankova, E., Torre, C. M., Rocha, A. M. A., Taniar, D., Apduhan, B. O., Tarantino, E., and Ryu, Y., editors, *Computational Science and Its Applications – ICCSA 2018*, pages 351–365. New York, USA: Springer International Publishing.

Bloch, P. H., Ridgway, N. M., and Dawson, S. A. (1994). The shopping mall as consumer habitat. *Journal of Retailing*, 70(1):23 – 42.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126 – 139.

Borst, H. C., Miedema, H. M., [de Vries], S. I., Graham, J. M., and [van Dongen], J. E. (2008). Relationships between street characteristics and perceived attractiveness for walking reported by elderly people. *Journal of Environmental Psychology*, 28(4):353 – 361.

Boulos, K. (2004). Towards evidence-based, gis-driven national spatial health information infrastructure and surveillance services in the united kingdom. *International journal of health geographics*, 3:1–50.

Boyd, D. and Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5):662–679.

Bradlow, E. T., Gangwar, M., Kopalle, P., and Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1):79 – 95.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

British Retail Consortium (2019). 21st century high streets: A new vision for our town centres [online; accessed 21 november 2019]. Available at: http://www.boots-uk.com/media/3866/21st_century_high_streets_2012-1-1.pdf.

Brito, P. Q., McGoldrick, P. J., and Raut, U. R. (2019). Shopping centre patronage: Situational factors against affect. *Vision*, 23(2):189–196.

Brown, S. (1992). *Retail Location: A Micro-scale Perspective*. Aldershot, UK: Avebury Publishing.

Browne, W. J., Subramanian, S. V., Jones, K., and Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(3):599–613.

Brueckner, J. K., Thisse, J.-F., and Zenou, Y. (1999). Why is central paris rich and downtown detroit poor?: An amenity-based theory. *European Economic Review*, 43(1):91 – 107.

CACI (2018). *Retail Footprint - CACI's leading retail centre catchment model*. CACI Intl. [Online; accessed 8 January 2020]. Available at: https://www.caci.co.uk/products/product/retail-footprint.

Calvert, G., Spence, C., and Stein, B. (2004). *The handbook of multisensory processing*. Cambridge, MA, USA: MIT Press.

Cao, X. J., Chen, Q., and Choo, S. (2013). Geographic distribution of e-shopping: Application of structural equation models in the twin cities of minnesota. *Transportation Research Record*, 2383(1):18–26.

Carmona, M. (2015). London's local high streets: The problems, potential and complexities of mixed street corridors. *Progress in Planning*, 100:1 – 84.

Carmona, M. (2019). Place value: place quality and its impact on health, social, economic and environmental outcomes. *Journal of Urban Design*, 24(1):1–48.

Cayo, M. R. and Talbot, T. O. (2003). Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, 2(10):1–12.

CDRC (2020). *Datasets*. Consumer Data Research Centre [Online; accessed 8 January 2020]. Available at: https://data.cdrc.ac.uk/search/type/dataset.

Cervero, R., Sarmiento, O. L., Jacoby, E., Gomez, L. F., and Neiman, A. (2009). Influences of built environments on walking and cycling: Lessons from bogotá. *International Journal of Sustainable Transportation*, 3(4):203–226.

Chalmers, K., Gessner, M., Venturoni, L., and Weiler, S. (2012). Stemming retail leakage with a sense of community: Leveraging the links between communal ties and shopping decisions. *The Social Science Journal*, 49(1):108 – 113.

Chebat, J.-C., Sirgy, M. J., and Grzeskowiak, S. (2010). How can shopping mall management best capture mall image? *Journal of Business Research*, 63(7):735 – 740.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754.

Chen, Y., Chen, W., Chen, Y., Tsai, B., Wang, Y. F., and Sun, M. (2017). No more discrimination: Cross city adaptation of road scene segmenters. *CoRR*, abs/1704.08509.

Cho, S. and Workman, J. (2011). Gender, fashion innovativeness and opinion leadership, and need for touch: Effects on multiâ€channel choice and touch/nonâ€touch preference in clothing shopping. *Journal of Fashion Marketing and Management: An International Journal*, 15(3):363–382.

Chollet, F. (2015). Keras [online; accessed 21 november 2019]. Available at: https://keras.io.

Christen, P. (2012). *Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*. New York, NY: Springer.

Christen, P. and Churches, T. (2019). Febrl data set generator program 'generate.py' [online; accessed 21 november 2019]. Available at: http://users.cecs.anu.edu.au/~Peter.Christen/Febrl/febrl-0.3/febrldoc-0.3/node70.html.

Chrysochou, P. (2017). *Consumer Behavior Research Methods*, pages 409–428. New York, USA: Springer International Publishing.

Churches, T., Christen, P., Lim, K., and Zhu, J. X. (2002). Preparation of name and address data for record linkage using hidden markov models. *BMC medical informatics and decision making*, 2:9–9.

Cibella, N. and Tuoto, T. (2012). Statistical perspective on blocking methods when linking large data-sets. *Advanced Statistical Methods for the Analysis of Large Data-Sets*, pages 81–89.

Clark, T. (2004). *The City as an Entertainment Machine*, chapter 3. Urban Amenities: Lakes, Operate and Juice Bars: Do they Drive Development?, pages 103–140.

Clark, T. and Chang, S. P. (2014). A history of tesco: the rise of britain's biggest supermarket [online; accessed 1 november 2019]. Available at: https://www.telegraph.co.uk/finance/markets/2788089/A-history-of-Tesco-The-rise-of-Britains-biggest-supermarket.html.

Clark, T. N., Lloyd, R., Wong, K. K., and Jain, P. (2002). Amenities drive urban growth. *Journal of Urban Affairs*, 24(5):493–515.

Clark, W. (1968). Consumer travel patterns and the concept of range. *Annals of the Association of American Geographers*, 58(2):386–396.

Clarke, G., Thompson, C., and Birkin, M. (2015). The emerging geography of e-commerce in british retailing. *Regional Studies, Regional Science*, 2(1):371–391.

Claycombe, R. J. (1991). Spatial retail markets. *International Journal of Industrial Organization*, 9(2):303 – 313.

Colquitt, J. and Zapata-Phelan, C. (2007). Trends in theory building and theory testing:

A five-decade study of the academy of management journal. *Academy of Management Journal*, 50(6):1281–1303.

Comber, S., Arribas-Bel, D., Singleton, A., Dong, G., and Dolega, L. (2019). Building hierarchies of retail centers using bayesian multilevel models. *Annals of the American Association of Geographers*, 0(0):1–24.

Coveney, P., Dougherty, E., and Highfield, R. (2016). Big data need big theory too. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374:37420160153.

Cyr, D. (2008). Modeling web site design across cultures: Relationships to trust, satisfaction, and e-loyalty. *Journal of Management Information Systems*, 24(4):47–72.

Daras, K., Davies, A., Green, M. A., and Singleton, A. (2018). Developing indicators for measuring health-related features of neighbourhoods. In Longley, P., Cheshire, J., and Singleton, A., editors, *Consumer Data Research*, chapter 12, pages 167–177. London, UK: UCL Press.

Davies, A., Dolega, L., and Arribas-Bel, D. (2019). Buy online collect in-store: exploring grocery click and collect using a national case study. *International Journal of Retail and Distribution Management*, 47(3):278–291.

Davies, R. L. (1973). *Patterns and profiles of consumer behaviour*. Newcastle, UK: University of Newcastle.

Dawson, J., Findlay, A., and Sparks, L. (2008). *The Retailing Reader*. Oxford, UK: Routledge.

Dawson, J. A. (2013). *Retail geography*. Oxford, UK: Routledge.

DCMS (2013). *Building Digital UK*. Department for Digital, Culture, Media and Sport

[Online; accessed 8 January 2020]. Available at: https://www.gov.uk/guidance/building-digital-uk.

de Palma, A., Dunkerley, F., and Proost, S. (2010). Trip chaining: who wins who loses? *Journal of Economics and Management Strategy*, 19:223–258.

Dekimpe, M. G. (2019). Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing*.

Dellaert, B. G. C., Arentze, T. A., Bierlaire, M., Borgers, A. W. J., and Timmermans, H. J. P. (1998). Investigating consumers' tendency to combine multiple shopping purposes and destinations. *Journal of Marketing Research*, 35(2):177–188.

dell'Olio, L., Ibeas, A., de Ona, J., and de Ona, R. (2018). Chapter 4 - designing a survey for public transport users. In dell'Olio, L., Ibeas, A., de Ona, J., and de Ona, R., editors, *Public Transportation Quality of Service*, pages 49 – 61. Amsterdam, NL: Elsevier.

Dennis, C., Marsland, D., and Cockett, T. (2002). Central place practice: shopping centre attractiveness measures, hinterland boundaries and the uk retail hierarchy. *Journal of Retailing and Consumer Services*, 9(4):185 – 199.

Dennis, C., Newman, A., Michon, R., Brakus, J. J., and Wright, L. T. (2010). The mediating effects of perception and emotion: Digital signage in mall atmospherics. *Journal of Retailing and Consumer Services*, 17(3):205 – 215.

Department for Communities and Local Government (2011). *Relaxation of planning rules for change of use from commercial to residential*. DCLG [Online; accessed 8 January 2021].

Department of the Environment Urban and Economic Development Group (1994). *High Street Performance and Evolution – A brief guide to the evidence*. Department of the Environment Urban and Economic Development Group [Online; accessed 26 March

2020]. Available at: http://www.nrpf.org.uk/PDF/Vital%20and%20Viable%20Town%20Centres%20-%20%20DoE%20(1994).pdf.

DfT (2014). *National Public Transport Access Nodes (NaPTAN)*. Department for Transport (DfT) [Online; accessed 8 August 2020]. Available at: https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan.

Dhar, S., Ordonez, V., and Berg, T. L. (2011). High level describable attributes for predicting aesthetics and interestingness. In *CVPR 2011*, pages 1657–1664.

Diesner, J. and Carley, K. M. (2008). Conditional random fields for entity extraction and ontological text coding. *Computational and Mathematical Organization Theory*, 14(3):248–262.

DiMaggio, P. J. (1995). Comments on "what theory is not". *Administrative Science Quarterly*, 40(3):391–397.

DoE (1996). *Town centres and retail development*. Department of the Environment. London: HMSO.

Doherty, N., Ellis-Chadwick, F., and Hart, C. (1999). Cyber retailing in the uk: the potential of the internet as a retail channel. *International Journal of Retail and Distribution Management*, 27(1):22–36.

Doherty, N. F. and Ellis-Chadwick, F. (2010). Internet retailing: the past, the present and the future. *International Journal of Retail & Distribution Management*, 38(11/12):943–965.

Dolega, L. and Lord, A. (2020). Exploring the geography of retail success and decline: A case study of the liverpool city region. *Cities*, 96:102456.

Dolega, L., Pavlis, M., and Singleton, A. (2016). Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services*, 28:78 – 90.

Dolega, L., Reynolds, J., Singleton, A., and Pavlis, M. (2019). Beyond retail: New ways of classifying uk shopping and consumption spaces. *Environment and Planning B: Urban Analytics and City Science*, 0(0):2399808319840666.

Dong, G. and Harris, R. (2015). Spatial autoregressive models for geographically hierarchical data structures. *Geographical Analysis*, 47(2):173–191.

Dong, G., Harris, R., Jones, K., and Yu, J. (2015). Multilevel modelling with spatial interaction effects with application to an emerging land market in beijing, china. *PLOS ONE*, 10(6):1–18.

Dong, G. and Wu, W. (2016). Schools, land markets and spatial effects. *Land Use Policy*, 59:366 – 374.

Donovan, R. J., Rossiter, J. R., Marcoolyn, G., and Nesdale, A. (1994). Store atmosphere and purchasing behavior. *Journal of Retailing*, 70(3):283 – 294.

Downs, R. M. (1970). The cognitive structure of an urban shopping center. *Environment and Behavior*, 2(1):13–39.

Du, H., Alechina, N., Jackson, M., and Hart, G. (2017). A method for matching crowdsourced and authoritative geospatial data. *Transactions in GIS*, 21(2):406–427.

Dubey, A., Naik, N., Parikh, D., Raskar, R., and Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision – ECCV 2016*, pages 196–212. New York, US: Springer International Publishing.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press, New York, NY, USA, 1st edition.

Ektefa, M., Sidi, F., Ibrahim, H., Jabar, M., and Memar, S. (2011). A comparative study in classification techniques for unsupervised record linkage model. *Journal of Computer Science*, 7:341–347.

El–Adly, M. I. (2007). Shopping malls attractiveness: a segmentation approach. *International Journal of Retail and Distribution Management*, 35(11):936–950.

Eppli, M. and Shilling, J. (1996). How critical is a good location to a regional shopping center? *Journal of Real Estate Research*, 12(3):459–468.

Ethical Consumer Research Association (2013). Ethical consumer market report [online; accessed 21 november 2019]. Available at: https://www.ethicalconsumer.org/sites/default/files/inline-files/ethical-consumer-markets-report-2013.pdf.

Finn, A. and Louviere, J. J. (1996). Shopping center image, consideration, and choice: Anchor store contribution. *Journal of Business Research*, 35(3):241 – 251.

Florida, R. (2000). *Competing in the age of talent: Quality of place and the new economy*. Richard King Mellon Foundation [Online; accessed 8 January 2020]. Available at: https://www.creativeclass.com/rfcgdb/articles/2000-Competing_In_The_Age_Of_Talent.pdf.

Florida, R. (2008). *Who's Your City?: How the Creative Economy Is Making Where You Live the Most Important Decision of Your Life*. Random House of Canada, Limited.

Fotheringham, S. (1983). A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning A: Economy and Space*, 15(1):15–36.

Foursquare (2020). Foursquare for developers – places data [online; accessed 21 april 2019]. Available at: https://developer.foursquare.com/places.

Garlin, F. and Owen, K. (2006). Setting the tone with the tune: A meta-analytic review of the effects of background music in retail settings. *Journal of Business Research*, 59(6):755–764.

Gatzlaff, D. H., Sirmans, G. S., and Diskin, B. A. (1994). The effect of anchor tenant loss on shopping center rents. *The Journal of Real Estate Research*, 9(1):99–110.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., and Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50):13108–13113.

Gehl, J. (2010). *Cities for People*. Washington DC, USA: Island Press.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*, volume Analytical methods for social research. Cambridge, UK: Cambridge University Press.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741.

Geofabrik (2020). Openstreetmap data for great britain [online; accessed 21 november 2019]. Available at: https://download.geofabrik.de/europe/great-britain/.

Ghose, A. and Han, S. (2014). Estimating demand for mobile applications in the new economy. *Management Science*, 60(6):1470–1488.

Girshick, R. (2015). Fast r-cnn. *arXiv preprint arXiv:1504.08083*.

Glaeser, E. L. and Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature*, 47(4):983–1028.

Glaeser, E. L., Kolko, J., and Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1):27–50.

Goldberg, D. W. (2011). Improving geocoding match rates with spatially-varying block metrics. *Transactions in GIS*, 15(6):829–850.

Goldman, W. P. and Seamon, J. G. (1992). Very long-term memory for odors: Retention of odor-name associations. *The American Journal of Psychology*, 105(4):549–563.

Goldstein, H. (2003). *Multilevel Modelling of Educational Data*, pages 25–42. Dordrecht, the Netherlands: Springer Netherlands.

Goldstein, H. (2011). *Multilevel statistical models*. West Sussex, UK: John Wiley & Sons.

Golledge, R. G., Rushton, G., and Clark, W. A. V. (1966). Some spatial characteristics of iowa's dispersed farm population and their implications for the grouping of central place functions. *Economic Geography*, 42(3):261–272.

Gomes, R. M. and Paula, F. (2017). Shopping mall image: systematic review of 40 years of research. *The International Review of Retail, Distribution and Consumer Research*, 27(1):1–27.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA, USA: MIT Press.

Google (2020). See how your community is moving around differently due to covid-19 [online; accessed 21 march 2019]. Available at: https://www.google.com/covid19/mobility/.

Goss, J. (1993). The "magic of the mall": An analysis of form, function, and meaning in the contemporary retail built environment. *Annals of the Association of American Geographers*, 83(1):18–47.

Grewal, D., Ahlbom, C.-P., Beitelspacher, L., Noble, S. M., and Nordfält, J. (2018a). In-store mobile phone use and customer shopping behavior: Evidence from the field. *Journal of Marketing*, 82(4):102–126.

Grewal, D., Motyka, S., and Levy, M. (2018b). The evolution and future of retailing and retailing education. *Journal of Marketing Education*, 40(1):85–93.

Grewal, D., Roggeveen, A. L., and Nordfält, J. (2017). The future of retailing. *Journal of Retailing*, 93(1):1 – 6.

Guy, C. (2007). *Planning for retail development. A critical view of the British experience.* London, UK: Routledge.

Hackett, P. M. W. and Foxall, G. R. (1994). A factor analytic study of consumers' location specific values: A traditional high street and a modern shopping mall. *Journal of Marketing Management*, 10(1-3):163–178.

Hallsworth, A. G. and Coca-Stefaniak, J. A. (2018). National high street retail and town centre policy at a cross roads in england and wales. *Cities*, 79:134–140.

Hara, K., Le, V., and Froehlich, J. (2013). Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, page 631–640, New York, NY, USA. Association for Computing Machinery.

Hardin, William G, I. and Wolverton, M. L. (2001). Neighborhood Center Image and Rents. *The Journal of Real Estate Finance and Economics*, 23(1):31–46.

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3):146–162.

Hart, C., Farrell, A. M., Stachow, G., Reed, G., and Cadogan, J. W. (2007). Enjoyment of the shopping experience: Impact on customers' repatronage intentions and gender influence. *The Service Industries Journal*, 27(5):583–604.

Hastie, T., Tibshirani, R., and Friendman, J. (2009). Random forests. In Hastie, T., Tibshirani, R., and Friendman, J., editors, *The elements of statistical learning: Data mining, inference and prediction*, chapter 10, pages 587–604. Springer, Berlin, Germany.

Hauser, J. R. and Koppelman, F. S. (1979). Alternative perceptual mapping techniques: Relative accuracy and usefulness. *Journal of Marketing Research*, 16(4):495–506.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. *arXiv preprint arXiv:1703.06870*.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.

Heffernan, E., Heffernan, T., and Pan, W. (2014). The relationship between the quality of active frontages and public perceptions of public spaces. *Urban Design International*, 19(1):92–102.

Helbich, M., Yao, Y., Liu, Y., Zhang, J., Liu, P., and Wang, R. (2019). Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in beijing, china. *Environment International*, 126:107 – 117.

Helm, S., Kim, S. H., and Van Riper, S. (2020). Navigating the 'retail apocalypse': A framework of consumer evaluations of the new retail landscape. *Journal of Retailing and Consumer Services*, 54.

Helsper, E. J. and Eynon, R. (2010). Digital natives: where is the evidence? *British Educational Research Journal*, 36(3):503–520.

Herbrich, R., Minka, T., and Graepel, T. (2007). Trueskill™: A bayesian skill rating system. In Schölkopf, B., Platt, J. C., and Hoffman, T., editors, *Advances in Neural Information Processing Systems 19*, pages 569–576. Cambridge, MA, USA: MIT Press.

Herzog, T., Scheuren, F., and Winkler, W. (2007). *Data Quality and Record Linkage Techniques*. New York, NY, USA: Springer.

Herzog, T. R. (1992). A cognitive analysis of preference for urban spaces. *Journal of Environmental Psychology*, 12(3):237 – 248.

Hilken, T., de ruyter, k., Chylinkski, M., Mahr, D., and Keeling, D. (2017). Augmenting the eye of the beholder: exploring the strategic potential of augmented reality to enhance online service experiences. *Journal of the Academy of Marketing Science*, 45:884–905.

Hillier, B. (1999). Centrality as a process: Accounting for attraction inequalities in deformed grids. *Urban Design International*, 4(3-4):107–127.

Hood, N., Urquhart, R., Newing, A., and Heppenstall, A. (2020). Sociodemographic and spatial disaggregation of e-commerce channel use in the grocery market in great britain. *Journal of Retailing and Consumer Services*, 55:102076.

Horowitz, J. L. and Louviere, J. J. (1995). What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, 12(1):39 – 54.

Huddleston, P. T., Behe, B. K., Driesener, C., and Minahan, S. (2018). Inside-outside: Using eye-tracking to investigate search-choice processes in the retail environment. *Journal of Retailing and Consumer Services*, 43:85 – 93.

Huff, D. and Jenks, G. (1968). A graphic interpretation of the friction of distance in gravity models. *Annals of the Association of American Geographers*, 58(4):814–824.

Huff, D. L. (1960). A topological model of consumer space preferences. *Papers in Regional Science*, 6(1):159–173.

Huff, D. L. (1963). A probabilistic analysis of shopping center trade areas. *Land Economics*, 39(1):81–90.

Huff, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing*, 28(3):34–38.

Hui, E., Yiu, C., and Yau, Y. (2007). Retail properties in hong kong: A rental analysis. *Journal of Property Investment and Finance*, 25(2):136–146.

Hui, E. C. and Liang, C. (2016). Spatial spillover effect of urban landscape views on property price. *Applied Geography*, 72:26 – 35.

Hwang, S.-S. and Albrecht, D. E. (1987). Constraints to the fulfillment of residential preferences among texas homebuyers. *Demography*, 24(1):61–76.

Hübner, A., Holzapfel, A., and Kuhn, H. (2016). Distribution systems in omni-channel retailing. *Business Research*, 9(2):255–296.

Ibrahim, M. (2002). Disaggregating the travel components in shopping centre choice: An agenda for valuation practices. *Journal of Property Investment and Finance*, 20:277–294.

Ibrahim, M. R., Haworth, J., and Cheng, T. (2020). Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*, 96:102481.

ICSC (2017). *The Socio-Economic Impact of European Retail Real Estate*. International Council of Shopping Centers [Online; accessed 28 January 2020]. Available at: https://www.icsc.com/uploads/t07-subpage/European_Impact_Study-2017.pdf.

IGD (2014). *Shopper Trends*. Institute of Grocery Distribution [Online; accessed 8 January 2020]. Available at: http://www.igd.com/our-expertise/Shopper-Insight/shopper-outlook/4893/Shopper-Trends.

IGD (2018). *Food and grocery market to grow 14.8 percent by £28.2bn by 2023*. Institute of Grocery Distribution [Online; accessed 8 January 2020]. Available at: https://www.igd.com/about-us/media/press-releases/press-release/t/uk-food-and-grocery-market-to-grow-148-by-282bn-by-2023/i/19052.

Ilic, L., Sawada, M., and Zarzelli, A. (2019). Deep mapping gentrification in a large canadian city using deep learning and google street view. *PLOS ONE*, 14(3):1–21.

Illeris, S. (1967). Functional regions in demark about 1960: Theoretical models and empirical observations. *Geografisk Tidsskrift*, 66:225–251.

Inkscape Project (2019). Amazon eyes gap as tube's 24 hour move signals the end for ticket office [online; accessed 21 november 2019]. Available at: https://inkscape.org.

Isard, W. (1956). *Location and space-economy.* Cambridge, MA, USA: MIT Press.

Jacobs, J. (1961). *The Death and Life of Great American Cities.* New York, USA: Random House.

Jalaladdini, S. and Oktay, D. (2012). Urban public spaces and vitality: A socio-spatial analysis in the streets of cypriot towns. *Procedia - Social and Behavioral Sciences*, 35:664 – 674.

Jaro, M. (1984). *Record linkage research and the calibration of record linkage algorithms (Statistical Research Division Report Series No. Census/SRD/RR-84/27).* U.S. Bureau of the Census, Washington, DC.

Jenks, G. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7:186–190.

Jensen-Butler, C. (1972). Nearest neighbour analysis of a central place system. *Tijdschrift voor economische en sociale geografie*, 63(5):353–359.

Jin, B. and Sternquist, B. (2004). Shopping is truly a joy. *The Service Industries Journal*, 24(6):1–18.

Johnston, R. J. (1966). The distribution of an intra-metropolitan central place theory. *Australian Geographical Studies*, 4(1):19–33.

Jones, C. and Livingstone, N. (2018). The 'online high street' or the high street online? the implications for the urban retail hierarchy. *The International Review of Retail, Distribution and Consumer Research*, 28(1):47–63.

Jones, K. (1991). Specifying and estimating multilevel models for geographical research. *Transactions of the Institute of British Geographers*, 16(2):148–159.

Joseph, L. and Kuby, M. (2011). *Gravity Modeling and its Impacts on Location Analysis*, pages 423–443. New York, USA: Springer US.

Juan, M. D. D. (2004). Why do people choose the shopping malls? the attraction theory revisited. *Journal of International Consumer Marketing*, 17(1):71–96.

Jung, H., young Lee, S., Kim, H. S., and Lee, J. S. (2017). Does improving the physical street environment create satisfactory and active streets? evidence from seoul's design street project. *Transportation Research Part D: Transport and Environment*, 50:269 – 279.

Jurek-Loughrey, A. and Deepak, P. (2019). *Semi-supervised and Unsupervised Approaches to Record Pairs Classification in Multi-Source Data Linkage*, pages 55–78. New York, USA: Springer International Publishing.

Kalliatakis, G. (2017). Keras-vgg16-places365 [online; accessed 21 november 2019]. Available at: https://github.com/GKalliatakis/Keras-VGG16-places365.

Kandt, J. and Batty, M. (2020). Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, page 102992.

Kannan, P. and Li, H. 2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1):22 – 45.

Kant, I. (1790). *Critique of Judgment*. New York, USA: Barnes & Noble.

Kelejian, H. H. and Prucha, I. R. (2007). Hac estimation in a spatial framework. *Journal of Econometrics*, 140(1):131 – 154.

Kelling, G. and Coles, C. (1997). *Fixing Broken Windows: Restoring Order And Reducing Crime In Our Communities*. New York, US: Simon and Schuster.

Kelly, C., Wilson, J., Baker, E., Miller, D., and Schootman, M. (2013). Using google street view to audit the built environment: Inter-rater reliability results. *Annals of behavioral medicine*, 45:108–112.

Kenyon, J. (1967). On the relationship between central place function and size of place. *Annals of the Association of American Geographers*, 57(4):736–750.

Kim, A. J. and Ko, E. (2012). Do social media marketing activities enhance customer equity? an empirical study of luxury fashion brand. *Journal of Business Research*, 65(10):1480 – 1486.

Kim, S. G., Cho, S.-H., Lambert, D. M., and Roberts, R. K. (2010). Measuring the value of air quality: application of the spatial hedonic model. *Air Quality, Atmosphere & Health*, 3(1):41–51.

Kim, Y.-K. (2002). Consumer value: an application to mall and internet shopping. *International Journal of Retail and Distribution Management*, 30(12):595–602.

Kitchin, R. (2016). The ethics of smart cities and urban science. *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences*, 374:20160115.

Koster, H. R., Pasidis, I., and van Ommeren, J. (2019). Shopping externalities and retail concentration: Evidence from dutch shopping streets. *Journal of Urban Economics*, 114:103194.

Kotzab, H. and Teller, C. (2003). Value-adding partnerships and co-opetition in the grocery industry. *International Journal of Physical Distribution and Logistics Management*, 33(3):268–281.

Krishna, A. (2013). *Customer sense: How the 5 senses influence buying behaviour*. New York, USA: Palgrave Macmillan.

Krishna, A., Lwin, M. O., and Morrin, M. (2010). Product Scent and Memory. *Journal of Consumer Research*, 37(1):57–67.

Kumar, V., Anand, A., and Song, H. (2017). Future of retailer profitability: An organizing framework. *Journal of Retailing*, 93(1):96 – 119.

Köpcke, H. and Rahm, E. (2010). Frameworks for entity matching: A comparison. *Data and Knowledge Engineering*, 69(2):197 – 210.

Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, page 282–289. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Lakshmanan, J. R. and Hansen, W. G. (1965). A retail market potential model. *Journal of the American Institute of Planners*, 31(2):134–143.

Lambiri, D., Faggian, A., and Wrigley, N. (2017). Linked-trip effects of 'town-centre-first' era foodstore development: An assessment using difference-in-differences. *Environment and Planning B: Urban Analytics and City Science*, 44(1):160–179.

Lamey, L., Deleersnyder, B., Steenkamp, J.-B. E., and Dekimpe, M. G. (2018). New product success in the consumer packaged goods industry: A shopper marketing approach. *International Journal of Research in Marketing*, 35(3):432 – 452.

Land Registry (2016). *Price Paid Open Data*. Office for National Statistics (ONS) [Online; accessed 8 January 2020]. Available at: http://landregistry.data.gov.uk/.

Law, S., Seresinhe, C., Shen, Y., and Gutiérrez-Roig, M. (2018). Street-frontage-net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, pages 1–27.

LDC (2020a). *About us*. Local Data Company (LDC) [Online; accessed 12 December 2019]. Available at: https://www.localdatacompany.com/company/about.

LDC (2020b). *Products – Online Analytics Platforms*. Local Data Company (LDC) [Online; accessed 28 January 2020]. Available at: https://www.localdatacompany.com/products.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature Cell Biology*, 521(7553):436–444.

Lehmann, D. R. (2020). The evolving world of research in marketing and the blending of theory and data. *International Journal of Research in Marketing*, 37(1):27–42.

Lemon, K. N. and Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6):69–96.

LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *The Review of Regional Studies*, 44(1):13–32.

Lesage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1-2):113–129.

Li, R., Lin, D. K., and Li, B. (2013). Statistical inference in massive data sets. *Applied Stochastic Models in Business and Industry*, 29(5):399–409.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., and Zhang, W. (2015). Assessing street-level urban greenery using google street view and a modified green view index. *Urban Forestry and Urban Greening*, 14(3):675 – 685.

Liu, L., Silva, E. A., Wu, C., and Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems*, 65:113 – 125.

Liu, L., Zhou, B., Zhao, J., and Ryan, B. D. (2016). C-image: city cognitive mapping through geo-tagged photos. *GeoJournal*, 81(6):817–861.

Llewellyn-Davies. (2007). *Urban Design Compendium*. London, UK: English Partnerships.

Lloyd, R. and Clark, T. (2001). The city as an entertainment machine. *Critical Perspectives of Urban Redevelopment*, 6(3):357–378.

Longley, P. A. and Singleton, A. D. (2009). Linking social deprivation and digital exclusion in england. *Urban Studies*, 46(7):1275–1298.

Lösch, A. (1940). *The economics of location*. New Haven, CT, USA: Yale University Press.

Louviere, J. J., Hensher, D. A., Swait, J. D., and Adamowicz, W. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge, UK: Cambridge University Press.

Louvière, J. and Timmermans, H. (1990). Stated preference and choice models applied to recreation research: A review. *Leisure Sciences*, 12(1):9–32.

Lu, Q. S. and Miller, R. (2019). How social media communications combine with customer loyalty management to boost green retail sales. *Journal of Interactive Marketing*, 46:87 – 100.

Lunt, P. and Livingstone, S. (1992). *Mass Consumption and Personal Identity: Everyday Economic Experience*. Maidenhead, UK: Open University Press.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA, USA: MIT Press.

Léo, P.-Y. and Philippe, J. (2002). Retail centres: Location and consumer's satisfaction. *The Service Industries Journal*, 22(1):122–146.

Madzharov, A. V., Block, L. G., and Morrin, M. (2015). The cool scent of power: Effects of ambient scent on consumer preferences and choice behavior. *Journal of Marketing*, 79(1):83–96.

Malone, T. W., Yates, J., and Benjamin, R. I. (1987). Electronic markets and electronic hierarchies. *Commun. ACM*, 30(6):484–497.

Martineau, P. (1958). The personality of a retail store. *Harvard Business Review*, 36:47–55.

Massey, D. (1991). A global sense of place. *Marxism Today*, 38:24–29.

Matthews, J. W. and Turnbull, G. K. (2007). Neighborhood street layout and property value: The interaction of accessibility and land use mix. *The Journal of Real Estate Finance and Economics*, 35(2):111–141.

Mazza, L. and Rydin, Y. (1997). Urban sustainability: Discourses, networks and policy tools. *Progress in Planning*, 47(1):1 – 74.

McConnell, K. E. and Strand, I. (1981). Measuring the cost of time in recreation demand analysis: An application to sportfishing. *American Journal of Agricultural Economics*, 63(1):153–156.

McFadden, D. (2001). Economic choices. *American Economic Review*, 91(3):351–378.

McInnes, B. T. and Pedersen, T. (2013). Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. *Journal of Biomedical Informatics*, 46(6):1116 – 1124.

Mejia, L. C. and Benjamin, J. D. (2002). What do we know about the determinants of shopping center sales? spatial vs. non-spatial factors. *Journal of Real Estate Literature*, 10(1):3–26.

Merleau-Ponty, M. (2004). *The World of Perception*. Oxford, UK: Routledge.

Michon, R., Chebat, J.-C., and Turley, L. (2005). Mall atmospherics: the interaction effects of the mall environment on shopping behavior. *Journal of Business Research*, 58(5):576 – 583.

Mikolov, T., Corrado, G., Chen, K., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, H. and Han, J. (2009). *Geographic Knowledge Discovery and Data Mining*. Boca Raton, FL, USA: CRC Press.

Miller, H. J. (2010). The data avalanche is here. shouldn't we be digging? *Journal of Regional Science*, 50(1):181–201.

Mitchell, B. (2014). *Workplace Zones: A new geography for workplace statistics*. Office for National Statistics (ONS) [Online; accessed 8 January 2020]. Available at: https://data.gov.uk/dataset/workplace-zones-a-new-geography-for-workplace-statistics3.

Moore, L. (1989). Modelling store choice: A segmented approach using stated preference analysis. *Transactions of the Institute of British Geographers*, 14(4):461–477.

Morikawa, T., Ben-Akiva, M., and McFadden, D. (2002). Discrete choice models incorporating revealed preferences and psychometric data. *Advances in Econometrics*, 16:29–55.

Mortimer, G., e Hasan, S. F., Andrews, L., and Martin, J. (2016). Online grocery shopping: the impact of shopping frequency on perceived risk. *The International Review of Retail, Distribution and Consumer Research*, 26(2):202–223.

Mumford, C., Parker, C., Ntounis, N., and Dargan, E. (2020). Footfall signatures and volumes: Towards a classification of uk centres. *Environment and Planning B: Urban Analytics and City Science*, 0(0):2399808320911412.

Murcio, R., Soundararaj, B., and Lugomer, K. (2018). *Movement in Cities: Footfall and its Spatio-Temporal Distribution*, pages 84–95. London, UK: UCL Press.

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., and Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29):7571–7576.

Nasar, J. L. (1987). Environmental correlates of evaluative appraisals of central business district scenes. *Landscape and Urban Planning*, 14:117 – 130.

Nase, I., Berry, J., and Adair, A. (2013). Hedonic modelling of high street retail properties: A quality design perspective. *Journal of Property Investment and Finance*, 31(2):160–178.

Nase, I., Berry, J., and Adair, A. (2015). Urban design quality and real estate value: in search of a methodological framework. *Journal of Urban Design*, 20(5):563–581.

Nasseh, D. and Stausberg, J. (2016). Evaluation of a binary semi-supervised classification technique for probabilistic record linkage. *Methods of information in medicine*, 55(2):136—143.

Nelson, R. (1958). *The selection of retail locations*. New York, USA: F.W. Dodge Corporation.

Newcombe, H. B. and Kennedy, J. M. (1962). Record linkage: Making maximum use of the discriminating power of identifying information. *Commun. ACM*, 5(11):563–566.

Nguyen, Q. C., Khanna, S., Dwivedi, P., Huang, D., Huang, Y., Tasdizen, T., Brunisholz, K. D., Li, F., Gorman, W., Nguyen, T. T., and Jiang, C. (2019). Using google street view to examine associations between built environment characteristics and u.s. health outcomes. *Preventive Medicine Reports*, 14:100859.

Nguyen, Q. C., Sajjadi, M., McCullough, M., Pham, M., Nguyen, T. T., Yu, W., Meng, H.-W., Wen, M., Li, F., Smith, K. R., Brunisholz, K., and Tasdizen, T. (2018). Neighbourhood looking glass: 360º automated characterisation of the built environment for neighbourhood effects research. *Journal of Epidemiology & Community Health*, 72(3):260–266.

Nikhashemi, S., Jebarajakirthy, C., and Nusair, K. (2019). Uncovering the roles of retail brand experience and brand love in the apparel industry: Non-linear structural equation modelling approach. *Journal of Retailing and Consumer Services*, 48:122 – 135.

Oakes, S. (2003). Musical tempo and waiting perceptions. *Psychology & Marketing*, 20(8):685–705.

O'Brien, L. and Harris, F. (1991). *Retailing: Shopping, Society, Space.* London, UK: David Fulton.

Odell, M. and Pickford, J. (2013). Amazon eyes gap as tube's 24 hour move signals the end for ticket office [online; accessed 21 november 2019]. Available at: https://www.ft.com/content/bc5024ce-528c-11e3-8586-00144feabdc0.

ODPM (2004). *Safer Places: The planning system and crime prevention.* Office of the Deputy Prime Minister (ODPM). [Online; accessed 8 January 2020]. Available at: https://www.gov.uk/government/publications/safer-places-the-planning-system-and-crime-prevention.

Oliveira, G., Bierrenbach, A., Camargo, K., Coeli, C., and Pinheiro, R. (2016). Accuracy of probabilistic and deterministic record linkage: the case of tuberculosis. *Revista de Saúde Pública*, 50:1–9.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, USA: Crown Publishing Group.

Oner, O. (2017). Retail city: the relationship between place attractiveness and accessibility to shops. *Spatial Economic Analysis*, 12(1):72–91.

ONS (2016). *National Records of Scotland ; Northern Ireland Statistics and Research Agency (2016): 2011 Census aggregate data. UK Data Service (Edition: June 2016)*. Office for National Statistics (ONS) [Online; accessed 8 January 2020]. Available at: http://dx.doi.org/10.5257/census/aggregate-2011-1.

ONS (2018). *Internet access – households and individuals, Great Britain Statistical bulletins*. Office for National Statistics (ONS) [Online; accessed 8 January 2020]. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018.

ONS (2019). *Introduction to Output Areas - the building block of Census geography*. Office for National Statistics (ONS) [Online; accessed 8 January 2020]. Available at: https://www.ons.gov.uk/census/2001censusandearlier/dataandproducts/outputgeography/outputareas.

OpenCage (2018). Opencage address formatting [online; accessed 21 november 2019]. Available at: https://github.com/OpenCageData/address-formatting.

OpenStreetMap (2018). Osmosis [online; accessed 21 november 2019]. Available at: https://github.com/openstreetmap/osmosis.

Oppewal, H. and Holyoake, B. (2004). Bundling and retail agglomeration effects on shopping behavior. *Journal of Retailing and Consumer Services*, 11(2):61 – 74.

Oppewal, H., Timmermans, H. J. P., and Louviere, J. J. (1997). Modelling the effects of shopping centre size and store variety on consumer choice behaviour. *Environment and Planning A: Economy and Space*, 29(6):1073–1090.

Owen, G., Harris, R., and Jones, K. (2016). Under examination: Multilevel models, geography and health research. *Progress in Human Geography*, 40(3):394–412.

Ozuduru, B. and Guldmann, J.-M. (2014). Retail location and urban resilience: towards a new framework for retail policy. *S.A.P.I.E.N.S*, 6(1):1–13.

PAF (2018). *Postcode Address File: UK address database*. Royal Mail [Online; accessed 8 January 2019]. Available at: https://www.poweredbypaf.com/.

Page, S. J. and Hardyman, R. (1996). Place marketing and town centre management: A new tool for urban revitalization. *Cities*, 13(3):153 – 164.

Pahl, R. (1970). *Whose City?* London, UK: Penguin.

Parr, J. (2017). Central place theory: An evaluation. *Review of Urban & Regional Development Studies*, 29(3):151–164.

Parsons, A. (2009). Use of scent in a naturally odourless store. *International Journal of Retail and Distribution Management*, 37(5):440–452.

Pavlis, M., Dolega, L., and Singleton, A. (2018). A modified dbscan clustering method to estimate retail center extent. *Geographical Analysis*, 50(2):141–161.

Petermans, A., Kent, A., and Cleempoel, K. V. (2014). Photo-elicitation: Using photographs to read retail interiors through consumers' eyes. *Journal of Business Research*, 67(11):2243 – 2249.

Pirie, G. H. (1976). Thoughts on revealed preference and spatial behaviour. *Environment and Planning A: Economy and Space*, 8(8):947–955.

Ploeger, S. K., Sawada, M., Elsabbagh, A., Saatcioglu, M., Nastev, M., and Rosetti, E. (2016). Urban rat: New tool for virtual and site-specific mobile rapid data collection for seismic risk assessment. *Journal of Computing in Civil Engineering*, 30(2):04015006.

Poushneh, A. and Vasquez-Parraga, A. Z. (2017). Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. *Journal of Retailing and Consumer Services*, 34:229 – 234.

Pred, A. (1966). *The spatial dynamics of U.S. Urban-industrial growth 1800-1914: Interpretive and theoretical essays*. Cambridge, Mass, US: Oxford University Press.

Puccinelli, N. M., Chandrashekaran, R., Grewal, D., and Suri, R. (2013). Are men seduced by red? the effect of red versus black prices on price perceptions. *Journal of Retailing*, 89(2):115 – 125.

Puccinelli, N. M., Goodstein, R. C., Grewal, D., Price, R., Raghubir, P., and Stewart, D. (2009). Customer experience management in retailing: Understanding the buying process. *Journal of Retailing*, 85(1):15 – 30.

Quartier, K. and Vanrie, J. (2019). 28 - from photographs to real stores: Context squared. In Meiselman, H. L., editor, *Context*, pages 585 – 603. Cambridge, UK: Woodhead Publishing.

Quercia, D., O'Hare, N. K., and Cramer, H. (2014). Aesthetic capital: What makes london look beautiful, quiet, and happy? In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '14, page 945–955, New York, NY, USA. Association for Computing Machinery.

Rappaport, J. (2009). The increasing importance of quality of life. *Journal of Economic Geography*, 9:779–804.

Rathore, M. M., Paul, A., Hong, W.-H., Seo, H., Awan, I., and Saeed, S. (2018). Exploiting iot and big data analytics: Defining smart digital city using real-time urban data. *Sustainable Cities and Society*, 40:600 – 610.

Ravenscroft, N. (2000). The vitality and viability of town centres. *Urban Studies*, 37(13):2533–2549.

Rayburn, S. W. and Voss, K. E. (2013). A model of consumer's retail atmosphere perceptions. *Journal of Retailing and Consumer Services*, 20(4):400 – 407.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/publication/884893/en.

Reilly, W. (1931). *The law of retail gravitation*. New York, USA: Knickerbocker Press.

Reimers, V. and Clulow, V. (2004). Retail concentration: a comparison of spatial convenience in shopping strips and shopping centres. *Journal of Retailing and Consumer Services*, 11(4):207 – 221.

Reimers, V. and Clulow, V. (2009). Retail centres: it's time to make them convenient. *International Journal of Retail and Distribution Management*, 37(1):541–562.

Reinartz, W., Dellaert, B., Krafft, M., Kumar, V., and Varadarajan, R. (2011). Retailing innovations in a globalizing retail market environment. *Journal of Retailing*, 87:53 – 66.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks.

RetailEconomics (2020). *UK Retail Stats and Facts*. Office for National Statistics (ONS) [Online; accessed 8 January 2020]. Available at: https://www.retaileconomics.co.uk/library-retail-stats-and-facts.

Reynolds, P., Behren, J., Gunier, R., Goldberg, D., Hertz, A., and Smith, D. (2003). Childhood cancer incidence rates and hazardous air pollutants in california: An exploratory analysis. *Environmental health perspectives*, 111(4):663–668.

Rhodes, C. and Brien, P. (2017). *The retail industry: statistics and policy*. House of Commons Library [Online; accessed 8 January 2020]. Available at: http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06186.

Richardson, H. W. (1977). On the possibility of positive rent gradients. *Journal of Urban Economics*, 4(1):60 – 68.

Rivera-Batiz, F. L. (1988). Increasing returns, monopolistic competition, and agglomeration economies in consumption and production. *Regional Science and Urban Economics*, 18(1):125 – 153.

Roggeveen, A. L., Grewal, D., and Schweiger, E. B. (2020). The dast framework for retail atmospherics: The impact of in- and out-of-store retail journey touchpoints on the customer experience. *Journal of Retailing*, 96(1):128 – 137.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Rosiers, F. D., Theriault, M., and Menetrier, L. (2005). Spatial Versus Non-Spatial Determinants of Shopping Center Rents: Modeling Location and Neighborhood-Related Factors. *Journal of Real Estate Research*, 27(3):293–320.

Rowlingson, K. and McKay, S. (2011). *Wealth and the wealthy: Exploring and tackling inequalities between rich and poor*. Policy Press at the University of Bristol.

Rubenstein, H. and Goodenough, J. (1965). Contextual correlates of synonymy. *Communications of the ACM*, 8:627–633.

Ruiz, J.-P., Chebat, J.-C., and Hansen, P. (2004). Another trip to the mall: a segmentation study of customers based on their activities. *Journal of Retailing and Consumer Services*, 11(6):333 – 350.

Rundle, A., Bader, M., and Richards, C. (2011). Using google street view to audit neighborhood environments. *American Journal of Preventative Medicine*, 40(1):94–100.

Rushton, G. (1969). Analysis of spatial behaviour by revealed space preference. *Annals of the Association of American Geographers*, 59(2):391–400.

Rust, R. T. (2019). The future of marketing. *International Journal of Research in Marketing*.

Ryan-Collins, J., Cox, J., Potts, R., and Squires, P. (2010). *Re-imagining the high street - Escape from Clone Town Britain*. New Economics Foundation [Online; accessed 8 April 2020]. Available at: https://b.3cdn.net/nefoundation/1da089b4b1e66ba2b3_v8m6b0c0w.pdf.

Sabanoglu, T. (2020). *Retail sales volume percentage change in Great Britain 2010-2019*. Statista [Online; accessed 8 January 2020]. Available at: https://www.statista.com/statistics/286805/percentage-increase-in-retail-sales-volumes-in-great-britain/.

Sack, R. D. (1988). The consumer's world: Place as context. *Annals of the Association of American Geographers*, 78(4):642–664.

Salesses, P., Schechtner, K., and Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. *PLOS ONE*, 8(7):1–12.

Salleh, K. and Ruddock, L. (1999). The development of a hedonic price model of retail property facilities design. 15th Annual ARCOM Conference, pages 337–44, Liverpool John Moores University. Association of Researchers in Construction Management.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2):488–493.

Schlaepfer, F. and Fischhoff, B. (2010). When Are Preferences Consistent? The Effects of Task Familiarity and Contextual Cues on Revealed and Stated Preferences. SOI - Working Papers 1007, Socioeconomic Institute - University of Zurich.

Schlosser, A. E., White, T. B., and Lloyd, S. M. (2006). Converting web site visitors into buyers: How web site investment increases consumer trusting beliefs and online purchase intentions. *Journal of Marketing*, 70(2):133–148.

Schulz, N. and Stahl, K. (1996). Do consumers search for the highest price? oligopoly equilibrium and monopoly optimum in differentiated products markets. *RAND Journal of Economics*, 27(3):542–562.

Scott, P. (2007). *Geography and Retailing.* New Jersey, USA: Routledge, 1 edition.

Seresinhe, C. I., Preis, T., and Moat, H. S. (2017). Using deep learning to quantify the beauty of outdoor places. *Royal Society Open Science*, 4(7).

Severin, V., Louviere, J. J., and Finn, A. (2001). The stability of retail shopping choices over time and across countries. *Journal of Retailing*, 77(2):185 – 202.

Sevtsuk, A. and Kalvo, R. (2018). Patronage of urban commercial clusters: A network-based extension of the huff model for balancing location and size. *Environment and Planning B: Urban Analytics and City Science*, 45(3):508–528.

Sheth, J. N., Newman, B. I., and Gross, B. L. (1991). Why we buy what we buy: A theory of consumption values. *Journal of Business Research*, 22(2):159 – 170.

Shogren, J. F. (2006). Valuation in the lab. *Environmental and Resource Economics*, 34(1):163–172.

Silver, D. and Clark, T. (2016). *Scenescapes: How Qualities of Place Shape Social Life.* Chicago, US: University of Chicago Press.

Silver, N. (2012). *The Signal and the Noise: The Art and Science of Prediction.* London, UK: Penguin Books Limited.

Sinai, T. and Waldfogel, J. (2004). Geography and the internet: is the internet a substitute or a complement for cities? *Journal of Urban Economics*, 56(1):1 – 24.

Singleton, A. (2015). *Data challenges and rateable value adjustment.* Local Data Company (LDC) [Online; accessed 8 January 2019]. Available at: https://www.cdrc.ac.uk/case-study/data-challenges-and-rateable-value-adjustment/.

Singleton, A. and Arribas-Bel, D. (2019). Geographic data science. *Geographical Analysis*, 0(0).

Singleton, A., Dolega, L., Riddlesden, D., and Longley, P. (2016). Measuring the spatial vulnerability of retail centres to online consumption through a framework of e-resilience. *Geoforum*, 69:5 – 18.

Sirmans, C. and Guidry, K. A. (1993). The determinants of shopping center rents. *Journal of Real Estate Research*, 8(1):107–116.

Small, M. L. and Feldman, J. (2012). *Ethnographic Evidence, Heterogeneity, and Neighbourhood Effects After Moving to Opportunity*, pages 57–77. Dordrecht, the Netherlands: Springer Netherlands.

Smith, H. (1937). *Retail distribution: a critical analysis.* London, UK: Oxford University Press.

Sorokin, A. and Forsyth, D. (2008). Utility data annotation with amazon mechanical turk. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8.

Spangenberg, E. R., Sprott, D. E., Grohmann, B., and Tracy, D. L. (2006). Gender-congruent ambient scent influences on approach and avoidance behaviors in a retail store. *Journal of Business Research*, 59(12):1281–1287.

Spence, C., Puccinelli, N. M., Grewal, D., and Roggeveen, A. L. (2014). Store atmospherics: A multisensory perspective. *Psychology & Marketing*, 31(7):472–488.

Spielman, S. E. and Singleton, A. (2015). Studying neighborhoods using uncertain data from the american community survey: A contextual approach. *Annals of the Association of American Geographers*, 105(5):1003–1025.

Steenbruggen, J., Borzacchiello, M. T., Nijkamp, P., and Scholten, H. (2013). Mobile phone data from gsm networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. *GeoJournal*, 78(2):223–243.

Stuart, A. and Miller, H. (2014). *Business Rates*. Institute for Fiscal Studies [Online; accessed 18 April 2018]. Available at: http://www.ifs.org.uk/budgets/gb2014/gb2014_ch11.pdf.

Stubbings, P., Rowe, F., and Arribas-Bel, D. (2019). A hierarchical urban forest index using street-level imagery and deep learning. *Remote Sensing*, 11(12):1395.

Suddaby, R. (2006). From the editors: What grounded theory is not. *Academy of Management Journal*, 49(4):633–642.

Teller, C. and Elms, J. (2010). Managing the attractiveness of evolved and created retail agglomerations formats. *Marketing Intelligence and Planning*, 28:25–45.

Teller, C. and Elms, J. R. (2012). Urban place marketing and retail agglomeration customers. *Journal of Marketing Management*, 28(5-6):546–567.

Teller, C., Kotzab, H., and Grant, D. (2006). The consumer direct services revolution in grocery retailing: An exploratory investigation. *Managing Service Quality*, 16(1):78–96.

Teller, C. and Reutterer, T. (2008). The evolving concept of retail attractiveness: What makes retail agglomerations attractive when customers shop at them? *Journal of Retailing and Consumer Services*, 15(3):127 – 143.

Teller, C., Reutterer, T., and Schnedlitz, P. (2008). Hedonic and utilitarian shopper types in evolved and created retail agglomerations. *The International Review of Retail, Distribution and Consumer Research*, 18(3):283–309.

Teller, C. and Schnedlitz, P. (2012). Drivers of agglomeration effects in retailing: The shopping mall tenant's perspective. *Journal of Marketing Management*, 28(9-10):1043–1061.

Thurstain-Goodwin, M. and Unwin, D. (2000). Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. *Transactions in GIS*, 4(4):305–317.

Timmermans, H. (2001). Spatial choice models. In Smelser, N. J. and Baltes, P. B., editors, *International Encyclopedia of the Social and Behavioral Sciences*, pages 14768 – 14771. Oxford, UK: Pergamon.

Timmermans, H. (2004). *Retail location and consumer spatial choice behavior*, pages 133–147. New York, USA: Springer.

Twitter (2020). Publish and manage tweets, and analyze tweet data [online; accessed 9 february 2020]. Available at: https://developer.twitter.com/en/products/tweets.

Ullman, E. (1941). A theory of location for cities. *American Journal of Sociology*, 46(6):853–864.

Ulrich, M., Jahnke, H., Langrock, R., Pesch, R., and Senge, R. (2019). Distributional regression for demand forecasting in e-grocery. *European Journal of Operational Research*.

van der Maaten, L. and Hinton, G. (2008). Contextual correlates of synonymy. *Journal of Machine Learning Research*, 9:2579–2605.

Vanhuele, M. and Drèze, X. (2002). Measuring the price knowledge shoppers bring to the store. *Journal of Marketing*, 66(4):72–85.

Verstraete, G., Aghezzaf, E.-H., and Desmet, B. (2019). A data-driven framework for predicting weather impact on high-volume low-margin retail products. *Journal of Retailing and Consumer Services*, 48:169 – 177.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269.

VOA (2014). *VOA rating list downloads.* Valuation Office Agency (VOA) [Online; accessed 21 January 2020]. Available at: https://www.gov.uk/guidance/how-non-domestic-property-including-plant-and-machinery-is-valued.

VOA (2018). *How non-domestic (business) properties are valued.* Valuation Office Agency (VOA) [Online; accessed 21 January 2020]. Available at: https://voaratinglists.blob.core.windows.net/html/rlidata.htm.

Vyt, D., Jara, M., and Cliquet, G. (2017). Grocery pickup creation of value: Customers' benefits vs. spatial dimension. *Journal of Retailing and Consumer Services*, 39:145 – 153.

Vyvere, Y. V. D. (1994). Stated preference decompositional modelling and residential choice. *Geoforum*, 25(2):189 – 202.

Wagner, B., Fillis, I., and Johansson, U. (2003). E-business and e-supply strategy in small and medium sized businesses (smes). *Supply Chain Management: An International Journal*, 8(4):343–354.

Wagner, W. (1975). Preference structures for grocery shopping. *Proceedings of the Association of American Geographers*, 7:257–259.

Wakefield, K. L. and Baker, J. (1998). Excitement at the mall: Determinants and effects on shopping response. *Journal of Retailing*, 74(4):515 – 539.

Wang, K., Zhao, Q., Lu, J., and Yu, T. (2015). K-profiles: A nonlinear clustering method for pattern detection in high dimensional data. *BioMed research international*, 2015:918954.

Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Ward, J., Bitner, M., and Barnes, J. (1992). Measuring the prototypicality and meaning of retail environments. *Journal of Retailing*, 68(2):194–220.

Wardman, M. (1988). A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy*, 22(1):71–91.

Warf, B. (2010). Central place theory. In Warf, B., editor, *Encyclopedia of geography*, pages 381–382. Oaks, CA, US: SAGE Publications.

Warnaby, G., Bennison, D., Davies, B. J., and Hughes, H. (2002). Marketing uk towns and cities as shopping destinations. *Journal of Marketing Management*, 18(9-10):877–904.

Warren, M. (2007). The digital vicious cycle: Links between social disadvantage and digital exclusion in rural areas. *Telecommunications Policy*, 31(6):374 – 388.

Weinberg, B., Davis, L., and Berger, P. (2013). Perspectives on big data. *Journal of Marketing Analytics*, 1:187–201.

Weltevreden, J. W. (2007). Substitution or complementarity? how the internet changes city centre shopping. *Journal of Retailing and Consumer Services*, 14(3):192 – 207.

Williams, D. E. (2009). The evolution of e-tailing. *The International Review of Retail, Distribution and Consumer Research*, 19(3):219–249.

Wilson, L. O., Weiss, A. M., and John, G. (1990). Unbundling of industrial systems. *Journal of Marketing Research*, 27(2):123–138.

Wolf, L., Anselin, L., Arribas-Bel, D., and Mobley, L. (2020). On spatial and platial dependence: examining shrinkage in spatially-dependent multi- level models. *Annals of the American Association of Geographers*.

Wrigley, N. and Dolega, L. (2011). Resilience, fragility, and adaptation: New evidence on the performance of uk high streets during global economic crisis and its policy implications. *Environment and Planning A: Economy and Space*, 43(10):2337–2363.

Wrigley, N. and Lambiri, D. (2014). *High Street Performance and Evolution – A brief guide to the evidence*. Economic and Social Research Council (ESRC) [Online; accessed 8 January 2020]. Available at: https://thegreatbritishhighstreet.co.uk/pdf/GBHS-HighStreetReport.pdf.

Wrigley, N., Lambiri, D., Astbury, G., Dolega, L., Hart, C., Reeves, C., Thurstain-Goodwin, M., and Wood, S. (2015). *British high streets: from crisis to recovery? A comprehensive review of the evidence*. Economic and Social Research Council (ESRC).

Wrigley, N., Lambiri, D., and Cudworth, K. (2009). Linked trips and town centre viability. *Town and country planning: the quarterly review of the Town and Country Planning Association*, 78(10):433–438.

Yancey, W. (2005). *Evaluating string comparator performance for record linkage (Research Report Series, Statistics 2005-05)*. U.S. Bureau of the Census, Washington, DC.

Yang, D., Liu, X., He, H., and Li, Y. (2019). Air-to-ground multimodal object detection algorithm based on feature association learning. *International Journal of Advanced Robotic Systems*, 16(3):1–9.

Yavas, U. and Babakus, E. (2009). Modeling patronage behavior: A tri-partite conceptualization. *Journal of Consumer Marketing*, 26:516–526.

Yin, L., Cheng, Q., Wang, Z., and Shao, Z. (2015). 'big data' for pedestrian volume: Exploring the use of google street view images for pedestrian counts. *Applied Geography*, 63:337 – 345.

Yin, L. and Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and google street view imagery. *Applied Geography*, 76:147 – 153.

Zentner, A., Smith, M., and Kaya, C. (2013). How video rental patterns change as consumers move online. *Management Science*, 59(11):2622–2634.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., and Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, 180:148 – 160.

Zoltan, J. and Masiero, L. (2012). The relation between push motivation and activity consumption at the destination within the framework of a destination card. *Journal of Destination Marketing and Management*, 1(1):84 – 93.

Appendices

A — Transactions in GIS Paper

DOI: 10.1111/tgis.12522

RESEARCH ARTICLE

WILEY Transactions in GIS

Machine learning innovations in address matching: A practical comparison of word2vec and CRFs

Sam Comber | Daniel Arribas-Bel

Department of Geography and Planning, University of Liverpool School of Environmental Sciences, Liverpool, UK

Correspondence
Sam Comber, Department of Geography and Planning, University of Liverpool School of Environmental Sciences, Roxby Building, 74 Bedford St, Liverpool L69 7ZT, UK.
Email: s.comber@liverpool.ac.uk

Funding information
Economic and Social Research Council

Abstract

Record linkage is a frequent obstacle to unlocking the benefits of integrated (spatial) data sources. In the absence of unique identifiers to directly join records, practitioners often rely on text-based approaches for resolving candidate pairs of records to a match. In geographic information science, spatial record linkage is a form of geocoding that pertains to the resolution of text-based linkage between pairs of addresses into matches and non-matches. These approaches link text-based address sequences, integrating sources of data that would otherwise remain in isolation. While recent innovations in machine learning have been introduced in the wider record linkage literature, there is significant potential to apply machine learning to the address matching sub-field of geographic information science. As a response, this paper introduces two recent developments in text-based machine learning—conditional random fields and word2vec—that have not been applied to address matching, evaluating their comparative strengths and drawbacks.

1 | INTRODUCTION

Address matching, the process of identifying pairs of records with a spatial footprint, is increasingly required for enriching data quality in wide-ranging, real-world applications. With government bodies, businesses and healthcare agencies drowning in an ever-increasing deluge of data, a competitive advantage exists in the analysis of integrated data sources as opposed to analyzing databases in isolation (Christen, 2012). Yet, in reality, most real-world databases are noisy, inconsistent and replete with missing values. These issues complicate the integration of data. In fact, the acquisition of matched addresses is often key to spatially enabling data used for visualization or spatial data mining projects (Boulos, 2004). In the address matching context, while geospatial matching is directed by linking the geometric representations of spatial objects (Du, Alechina, Jackson, & Hart, 2017), spatial record linkage focuses on resolving text-based linkages between addresses.[1]

In the absence of unique identifiers that enable direct linking of data in relational database management environments, practitioners have traditionally relied on mathematical linkage techniques broadly divided by deterministic or probabilistic principles (Churches, Christen, Lim, & Zhu, 2002). While deterministic matching consists of generating hand-crafted rule bases for classification developed from specialist domain knowledge (Oliveira, Bierrenbach, Camargo, Coeli, & Pinhero, 2016), probabilistic linkage incorporates the varying distributions of a record's attribute values into the assignment of different weights for each field comparison. Weight assignment is related to the frequencies of value occurrences, with stronger weights given to matches for attributes upon which matching is less likely (Blanchette, DeKoven, De, & Roberts, 2013). For address matching, field comparisons might include comparing the street names of an address pair, with more common street names penalized by a lower weighting factor. In this way, resolving text-based postal addresses to the same address is a form of *geocoding*, where the quality of the match rate is intrinsically tied to the quality of the underlying reference data layer (Goldberg, 2011). Traditionally, address matching has focused on the probabilistic linkage approaches developed by the US Census Bureau in the 1970s (Jaro, 1984).

More recently, record linkage has been permeated by advances in machine learning. In this article, we focus on introducing two particular innovations into the address matching workflow: conditional random fields (CRFs) and word (address) embeddings. Before classification into address matches, input data requires segmentation into feature columns (Churches et al., 2002). The segmentation of postal addresses into attribute columns representing street numbers, street names or zip codes, for example, has been traditionally undertaken using hidden Markov models (HMMs). HMMs use statistical induction to predict, from possible arrangements of hypothetical states, the most likely arrangement to have produced the address sequence, and to then label each state by an attribute field (Christen, 2012). For addresses, labels might identify whether the present state represents a street number or street name. A recent innovation for text segmentation tasks has been the use of trained CRFs (Lafferty, McCallum, & Pereira, 2001). While HMMs assume the labeling of text sequences is statistically independent of previous outputs, CRFs are conditional by nature, meaning they assume no independence between output labels. Given that real-world text sequences such as addresses are represented by interaction and dependencies between words (e.g., zip codes are related to city names), it is reasonable to assume CRFs will perform well on a number of real-world text segmentation tasks.

A second innovation relates to the construction of so-called "comparison vectors" that are used for classifying records into matches and non-matches. Comparison vectors are created for each candidate record pair and contain several attributes that describe the text similarity of each pair (Christen, 2012). Traditionally, comparison vectors have been generated using string similarity metrics that measure the text distance between two address fields (e.g., the string similarity between two street names such as "Baker Street" and "Bakery Road"). Recently, however, advances from the natural language processing community demonstrate methods that map whole words and sentences to vectors in a continuous vector space. *Word2vec* (Mikolov, Chen, Corrado, & Dean, 2013) is one such method that maps semantically and syntactically similar words to nearby points in a vector space, encoding many linguistic patterns and regularities contained within the text. Such methods rely on a theory of language called the distributional hypothesis which states that words appearing in the same context purport similar meaning (Zellig, 1954). In the address matching context, one might hypothesize the word vectors generated for two semantically and syntactically similar postal addresses may be correctly resolved to a match.

While recent advances in machine learning have become adopted in the wider record linkage literature (Ektefa, Sidi, Ibrahim, Jabar, & Memar, 2011; Kopcke & Rahm, 2010; Nasseh & Stausberg, 2016), the address matching sub-field of geographic information science holds significant potential for the application of machine learning. In this article we explore how these advances can be integrated into the address matching workflow. In particular, we empirically evaluate the performance of CRFs and word2vec in computing high-quality match rates between pairs of postal addresses. The remainder of this article is organized as follows. Section 2 introduces the data challenge. Section 3 motivates the methodology of the workflow. Section 4 presents the findings of the address matching methods applied. Section 5 concludes.

2 | DATA

Our comparison relies on a set of addresses previously matched by the Local Data Company (LDC) (Singleton, 2015). This provides a ground truth of address pairs with a known match status, which allows us to evaluate the performance of the linkage methods. These address pairs were obtained from a previous round of matching between non-domestic addresses of the LDC and Valuation Office Agency (VOA) databases. In particular, these address pairs reflect matches between LDC records of high-street shops and commercial addresses contained in the VOA 2010 rating list (VOA, 2017). A description of the address fields for the LDC and VOA addresses that were segmented by a method we introduce later, the conditional random fields, is introduced in Table 1, to familiarize the reader with components of a structured address string.

Crucially, the matched set of LDC to VOA addresses contains 110,742 pairs that resolve to the same address. This matched set is augmented with 934,150 synthetic non-matched pairs that are generated with the Freely Extensible Biomedical Record Linkage (FEBRL) (Christen & Churches, 2005) data set generator. This works by creating variants of the matched addresses with different error characteristics introduced to the data, meaning the models learn the representations of non-matched addresses (Christen, 2012). Thus, for each address field in Table 1, with the exception of zip code, we introduce error characteristics to the data. A demonstration of how the synthetic non-matches are generated is given in Table 2, where three examples of records from the LDC and VOA database are mutated with different error characteristics. In particular, we set the probability of a missing field as proportional to the number of missing fields in the matched addresses. Moreover, we set the maximum number of modifications per address field and per address string to one, also testing a scenario where we increase the number of modifications to nine later on. These modifications introduce a probability for a character in the address field to be randomly inserted, deleted, substituted or transposed. Importantly, the match status of these synthetic non-matches is always set to false, meaning our machine learning techniques learn the representations of non-matched addresses for highly nuanced cases. To prepare the data for segmentation, we append all address fields from the LDC and VOA data sets into a comma-separated address string (e.g., “Home Bargains, 28, Church Way, Bradford”) while keeping the zip code separate for reasons we explain immediately below.

TABLE 1 CRF parser label tags and descriptions identified for the LDC and VOA addresses

Tag	Description	Example
House	Venue or business name ascribed to the address	Automotive Solutions
Number	Street-facing building number or apartment number	43
Unit	A secondary unit designator that identifies an office, unit or apartment	4a
Level	Expression signifying a floor number	Ground Floor
Street	Identifying name given to a street	Paradise Street
Suburb	Unofficial neighborhood name	Ropewalks
City	Any human settlement such as a metropolis, city, town or village	Liverpool
District	Second-level administrative division	North West
State	First-level administrative division	England
Zip code	Postal code used for mail sorting	L3 5TB

Note. Tags are aligned to address fields of the OpenCage (2018) address formatting library.

TABLE 2 FEBRL data set generator (Christen & Churches, 2005) demonstration for example records from the LDC and VOA database

ID	House	Number	Unit	Level	Street	Suburb	City	District	State
1-original-LDC	Kwik Fit	67–69	–	–	Whiteladies Road	–	Bristol	South West	
1-duplicate-1	–	41	Unit A	–	White Avenue	–	Bristol	South West	England
1-duplicate-2	Fitness Quick	64–66	–	–	Whiteladies Road	–	Bristol	–	
2-original-VOA	Thomson	–	Unit 19c	–	Teeside Retail Park	Goodwood Square	Stockton-on-Tees	North East	–
2-duplicate-1	–	5	Unit 1a	–	Teeside Retail Park	Goodewood Square	Stockton-on-Tees	North East	
2-duplicate-2	–	–	Unit 5b	–	Teaside Retail Park	Goodwood Square	Stockton-on-Tees	North East	
3-original-VOA	Body Shapers	–	72c	First	Ridgeway Street	–	Plymouth	South West	–
3-duplicate-1	–	–	8a	Second	Ridgeway Street	–	Plymouth	South West	–
3-duplicate-2	Bdoy Shapres	–	72e	–	Longmore Road	–	Plymouth	South West	–

Our next step introduces *blocking* for each method to increase the computational tractability of the linkage task. On a Dell Precision Tower 7000 series with 60 GB RAM and multi-core processor, the computational cost for the comparison of each address in the LDC and VOA data sets is substantial. This is because the number of address comparisons without blocking is a function of the Cartesian product of both data sets, which has quadratic complexity $O(n^2)$. So, for example, if the LDC and VOA data sets both contained just 10^4 records, the linkage task requires 10^8 comparisons, which becomes computationally non-trivial. To remedy this, we introduce blocking to partition the set of all possible address comparisons between the LDC and VOA databases to within mutually exclusive blocks (Newcombe & Kennedy, 1962). Now, if we let b equal the number of blocks, we are left with n/b addresses per partition on average, which reduces the complexity to $O(n^2/b)$ (Christen, 2012). This means the linkage task becomes tractable even on low-performance machines, as the linkage within each partition can be processed sequentially or, alternatively, in parallel if the user has access to a multi-core machine. Therefore, in each of the following methods, we use the zip codes of postal addresses as a blocking key. This reduces the number of address pair comparisons to within 39,855 zip code "blocks," with the distributional characteristics of these partitions displayed in Figure 1. In our case, the near uniform frequency distribution of the zip code blocks and completeness of the zip code attribute means it is a sensible choice as a blocking key. Yet, for different address databases where the zip code column is replete with missing values, an alternative attribute should be considered as a blocking key, which is an empirical decision to be motivated by the characteristics of the databases' attribute columns.

One potential issue with using zip codes as a blocking key is the existence of typographic errors in the spelling. In our case, this is pertinent because while validation checks are employed by the LDC, the recording of commercial addresses is undertaken by teams of surveyors, and is therefore susceptible to human error. To account for this, we explore *sorted neighborhood blocking*. We sort together the LDC and VOA data sets using the zip code as a sorting key value while restricting address comparisons to records within a window of fixed dimension, $w = 5$. As the window slides over the sorted zip codes, the identification of matches and non-matches is restricted to candidate address pairs within this window of fixed size (Cibella & Tuoto, 2012). This means that the technique is highly sensitive to lexicographic order which, in our case, is advantageous because we create candidate address

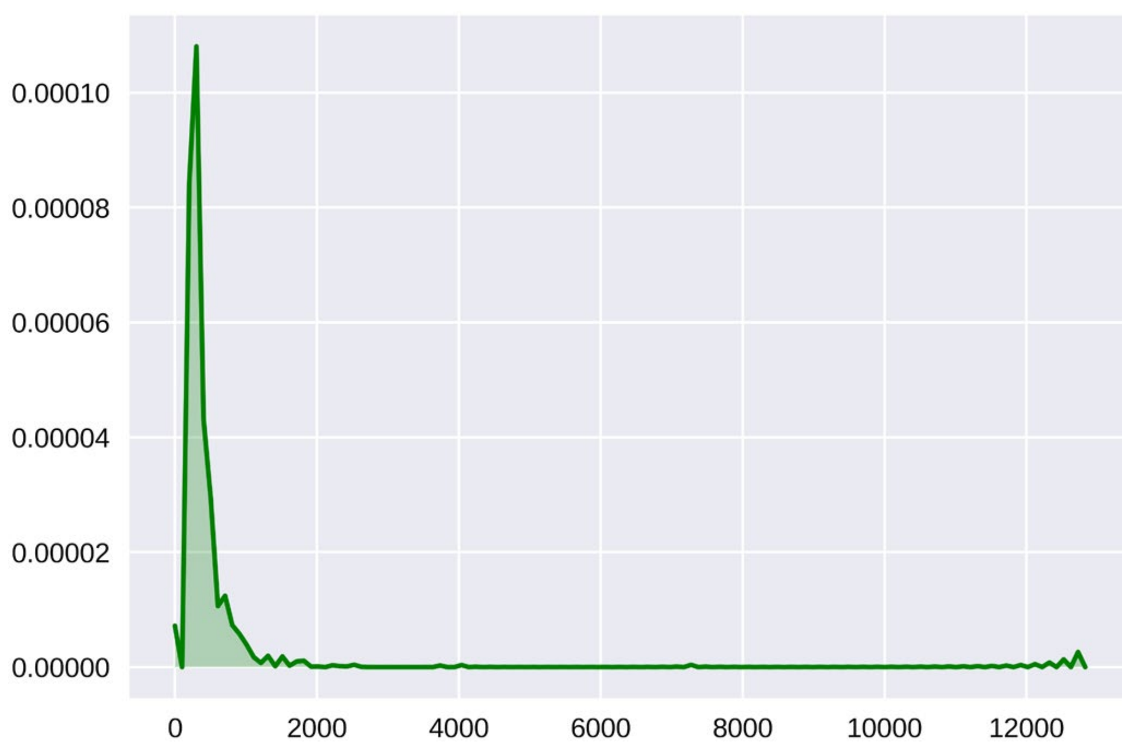


FIGURE 1 Distributional characteristics of the address block partitions (n = 39,855)

pairs from misspelt zip codes (e.g., the comparison between addresses with sorted zip codes "L3 5TA" and "L3 5TB" that pertain to the same address). However, a limitation of sorted neighborhood blocking is that misspellings of the first character in the blocking key can erroneously direct addresses to incorrect zip code blocks. *Q*-gram blocking is an alternative approach to generating partitions of zip code blocks, and converts blocking keys to a list of bigrams ($Q = 2$). The zip code "SW11 9LU", for example, generates a bigram list ["SW","W1","11","19","9L","LU"], from which k sub-lists of length $k-1$ are generated recursively to create variations of the zip code (Christen, 2012). This continues up to a minimum threshold for the number of bigrams in the sub-list. Following the example, this might create blocking key values such as "W1 9LU", "SW1 9LU" and "SW19 LU", meaning that the same address is likely to be inserted into several different zip code blocks. While the recursive generation of sub-lists is computationally expensive, the advantage of *Q*-gram blocking is that it overcomes typographic error in the characters of the blocking key. In our case, as the results presented below show, given the zip code attribute was profiled as generally being of high quality, the choice of blocking mechanism was less fundamental. Nevertheless, we introduce additional blocking techniques to provide instructional guidance to researchers interested in replicating our methodology on their own data sets, where attributes of the blocking key values might be less clean.

3 | METHODS

3.1 | Conditional random fields

One of the principal challenges in obtaining high-quality match rates is the conversion of raw data into a structured, usable format for comparison. For postal addresses, this involves parsing address sequences into feature columns. Take, for example, a canonical address of the form "3B Records, 5 Slater Street, Liverpool L1 4BW". Our objective is to segment (or label) this address into appropriate columns for business name, property number, street name, city and zip code. To use a hidden Markov model (Baum & Petrie, 1966) for segmentation would be to assign a joint probability to the observation sequence where the labeling of address elements is independent of previous labels (Churches et al., 2002). This means, following the example above, that "3B" could be incorrectly classified as a property number, whereas it actually completes the business name "3B Records". Importantly, "3B" will now be considered a property number, and that (alternative) fact will be used to classify the next token, "Records". This leads to an erroneous sequence of label predictions.

In real-world text sequences such as addresses, the probability of a transition between labels might depend not just on the current address element, but also on past and future elements. For this reason, conditional random fields (Lafferty et al., 2001) are more suited to addressing segmentation tasks. Principally, this is because CRFs negate what is known as the *label bias problem*: "transitions leaving a given state to compete only against each other, rather than against all transitions in the model" (Lafferty et al., 2001). When the CRFs has parsed "3B" and reaches the second token, "Records", the model scores an $l \times l$ matrix where l is the maximum number of labels that can be assigned by the model. In L, element l_{ij} reflects the score for the probability of the current word being labeled i, and the previous word labeled j (Diesner & Carley, 2008). Returning to the example, when the parser gets to the *actual* property number, "5", the highest score in the matrix indicates the current label should be revised to a property number, and the previous label to a business name. Below, we provide an illustrative example of an erroneous sequence of labels predictions that hypothetically may have been segmented with an HMM.

3B	*Records*	*Slater Street*	*Liverpool*	L1 4BW
NUMBER	STREET	SUBURB	CITY	ZIP CODE

In the CRFs, prediction of the most likely sequence of labels uses a reversible highest *scoring* path. This is known as Viterbi inference (Viterbi, 1967), and leads to a sequence of labels with the highest likelihood.

3B Records	*5*	*Slater Street*	*Liverpool*	L1 4BW
HOUSE	NUMBER	STREET	CITY	ZIP CODE

With the CRFs model, the previously raw and unstructured address will now be correctly segmented into the following feature columns that can be used as a basis for classifying records into matches and non-matches:

House:	3B Records
Number:	5
Street:	Slater Street
City:	Liverpool
Zip code:	L1 4BW

In our case, address segmentation is undertaken using the *Libpostal* C library (Barrentine, 2018) that trains a CRFs model on addresses sourced from OpenStreetMap (OSM) data. This means we apply a pre-trained address segmentation model to each raw address string, setting the country code of the parser to "GB" (for Great Britain) so the software recognizes it is segmenting UK addresses. Libpostal's *parse_address* command will then label the address sequence into features columns, if they exist, for the fields in Table 1. To empirically evaluate the performance gain, we also introduce an HMM[2] alongside the CRFs parser. Once feature columns consisting of address elements have been obtained for every LDC and VOA address, a *comparison vector* is constructed for each candidate address pair. Comparison vectors contain several attributes that describe the text similarity between each feature column (Christen, 2012). In our case, each element of this comparison vector contains the Jaro–Winkler string similarity between each address field, with exception of zip code, in Table 1. Briefly, the Jaro–Winkler distance calculates the minimum number of single character transpositions required to convert one string into another, also increasing the similarity when the first few characters are the same (Herzog, Scheuren, & Winkler, 2007). We motivate the decision to use Jaro–Winkler as our string comparison function because previous findings show it performs best on attributes containing named values (e.g., property names, street names, or city names) (Christen, 2012; Yancey, 2005). If a given address field is missing, the Jaro–Winkler similarity between the pair of address fields is set to zero. These comparison vectors for each address pair are the basis of a binary classification for classifying whether address pairs resolve to matches or non-matches. The general idea is that the more similar two addresses are, as described by Jaro–Winkler similarity, the higher the likelihood that they resolve to the same address.

Our classification approach is supervised, meaning we use our training data of known true match and true non-match status generated in Section 2 to evaluate the outcome of our address matching exercise. By training a classifier, we allow the model to learn the nuances of matched and non-matched addresses. This means that, after training, we can test whether unseen address pairs for which the match status is known correctly resolve to matches and non-matches, allowing us to evaluate the performance of our linkage techniques. Thus, once comparison vectors have been generated for each training record, we introduce several classifiers to facilitate the linkage into matches and non-matches. In particular, two ensemble methods for classification, a random forest (Breiman, 2001) and gradient boosted classifier known as XGBoost (Chen & Guestrin, 2016), are trained alongside a logistic regression model. We motivate our use of ensemble classifiers for one key reason. Our comparison vectors are embedded into a nine-dimensional vector space. This means each dimension reflects one of the nine Jaro–Winkler similarities between each address field in Table 1, with the exception of zip codes which are used for blocking. When partitioning matches from non-matches in this 9-dimensional vector space, while the logistic model searches for a linear decision boundary, the multiple decision trees of the ensemble methods partition the vector space into half spaces by using axis-aligned linear decision boundaries (Efron & Hastie, 2016). This has the net effect of a nonlinear decision boundary, which is desirable if the comparison vectors cannot be accurately separated into matches and non-matches by a single hyperplane. From here, the classifiers are trained using *k*-fold cross-validation where k = 10, which we explain in Section 4, and are evaluated against metrics commonplace in machine learning such as precision and recall (Christen, 2012).

3.2 | Word embeddings

In our second approach, we augment the use of CRFs with so-called "word embeddings," which is the name given to the vector representations of words. Vector space models *embed* words in a continuous vector space, where words with similar syntactic and semantic meaning are mapped, or embedded, to nearby points (Mikolov et al., 2013). Such methods leverage the distributional hypothesis of language which states that "words which are similar in meaning occur in similar contexts" (Rubenstein & Goodenough, 1965). One such method is word2vec (Mikolov et al., 2013), a neural probabilistic language model whose training objective is to find word vector representations that are good at predicting the surrounding words in a text-based sentence or document.[3] In our case, we hypothesize that learning high-dimensional vectors from postal addresses may be used to match addresses that resolve to the same geographic location despite irregularities in the text. In practice, we train *gensim*'s (Řehuřek & Sojka, 2010) implementation of word2vec on 29.6 million parsed postal addresses from the UK Postcode Address File (PAF) database (PAF, 2018). Learning word embeddings using word2vec requires setting the dimensionality of the vectors, so the training phase begins by randomly initializing each address field component with 100 real numbers. This means each parsed address field is represented by an array of numbers of length 100 which, as an example, can be represented as: $[0.32 \quad 0.28 \quad \ldots \quad 0.01 \quad 0.58] \in \mathcal{R}^{100}$. By feeding successive address fields into the model, the real numbers of each word vector are updated so that words sharing the same context are mapped closer together in the vector space. To build intuition for this idea, we employ Figure 2, where the t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten & Hinton, 2008) dimensionality reduction technique is applied to the top 10 closest vectors to an address

FIGURE 2 t-SNE visualization demonstrating the top 10 closest vectors to "Halifax PLC" in a two-dimensional vector space

field for property name, "halifax plc". In this two-dimensional vector space, the closeness of word vectors, measured by cosine similarity, represent words that share closer semantic and syntactic meaning. "Halifax PLC", for example, is a bank, and interestingly the word vector generated for it is embedded nearby to businesses that have a financial remit, "Natwest", "TSB Bank" and "Barclays Bank PLC", for example.

Under this approach, instead of using the Jaro–Winkler similarity between the address fields, we augment the linkage task by comparing word vectors generated by word2vec for the address fields segmented by the CRFs model. Thus, after training a word2vec model on PAF addresses, for every LDC and VOA address we are able to obtain a 100-dimensional vector for each address field. The postal address, "5 Myrtle Street, Liverpool", for example, contains three address fields (a street number, a street name, and a city name) for which we obtain vectors. Similarly to our first approach, we construct a comparison vector, but this time each element is the *cosine similarity* between the word vectors constructed from the address fields parsed by the CRFs model. In cases where address fields are missing, we set the cosine similarity to zero, which implies orthogonality or linear independence between the vectors under comparison. The decision to choose cosine similarity as the criterion for measuring similarity between address fields is because it has favorable qualities in capturing the semantic closeness of word vectors (McInnes & Pedersen, 2013). As before, we train a random forest, XGBoost and logistic regression model on the comparison vectors and associated match status labels using *k*-fold cross-validation to evaluate the linkage performance.

4 | RESULTS

To evaluate the performance of the HMM and CRFs alongside the CRFs augmentation with word2vec, we use address pairs for which the match status is known (as discussed in Section 2). The results of these methods are highlighted in Table 3, and are benchmarked by evaluation metrics known as *recall* and *precision*. Recall measures the proportion of address pairs that should have been classified, or recalled, as matched (Christen, 2012). The precision (or, equivalently, the positive predictive value) calculates the proportion of the matched address pairs that are classified correctly as true matches (Christen, 2012). To minimize over-fitting our supervised models, we introduce *k*-fold cross-validation, where *k*=10, meaning the training data is split into ten disjoint groups. In each split, the classifier is trained and tested on these subsets of address pairs, with the resulting recall and precision averaged across the groups. This means that for each group we have a randomized training and testing set split by 75% and 25%, respectively.

TABLE 3 Recall and precision evaluation metrics for the HMM, CRFs, and CRFs augmented using word2vec

Method	Precision	Recall
HMM		
Logistic	0.738	0.459
Random forest	0.944	0.696
XGBoost	0.959	0.688
CRFs		
Logistic	0.933	0.820
Random forest	0.940	0.918
XGBoost	0.955	0.902
CRFs–word2vec		
Logistic	0.870	0.687
Random forest	0.933	0.874
XGBoost	0.950	0.870

Note. Results are 10-fold cross-validated using 25% of the data for testing within each fold.

To begin interpretation, we first turn our attention to the baseline HMM that we use as a point of comparison for the machine learning techniques we introduce. Consistent with our earlier motivations, when address fields are parsed with the CRFs model, they outperform the HMM. This is shown by the lower recall values retrieved by each of the classifiers using the HMM technique. Interestingly, the precision values of the HMM and CRFs techniques are broadly consistent. This implies that, of the total number of matches returned, both techniques perform well at partitioning true positives from false positives, but the CRFs classify a larger proportion of matches, as shown by the higher recall value. This finding suggests that, unlike the HMM, the reversible sequence of labeling introduced by the CRFs leads to higher-quality match rates. We now turn to the supervised classifiers that are trained on the comparison vectors built using the CRFs model. An interesting facet of tree-based models is that the feature importances can be recovered (Hastie, Tibshirani, & Friedman, 2009). The "importance" of different features, or, equivalently, address fields, to the match classification is visualized by the red bars in Figure 3, along with the inter-trees variability. Figure 3 is consistent with conventional wisdom, as it indicates that street name and house number are the most important features that are used when resolving candidate pairs of addresses to a match. To visualize the absolute numbers of matches, we provide a confusion matrix in Figure 4 for the random forest trained with address pairs parsed by the CRFs. In Figure 4, the top left quadrant shows true negatives, top right shows false positives, bottom left shows false negatives, and bottom right shows true positives. Briefly, true positives are address pairs labeled as matches that are true matches; false positives are address pairs mislabeled as matches; true negatives are records classified as non-matches which are true non-matches; and false negatives are addresses classified as non-matches but are actually true matches (Christen, 2012).

From Table 3 it is clear the ensemble learners offer slight improvement over the logistic model in returning a larger fraction of true positives among all returned "matches". This is shown by the marginally higher precision

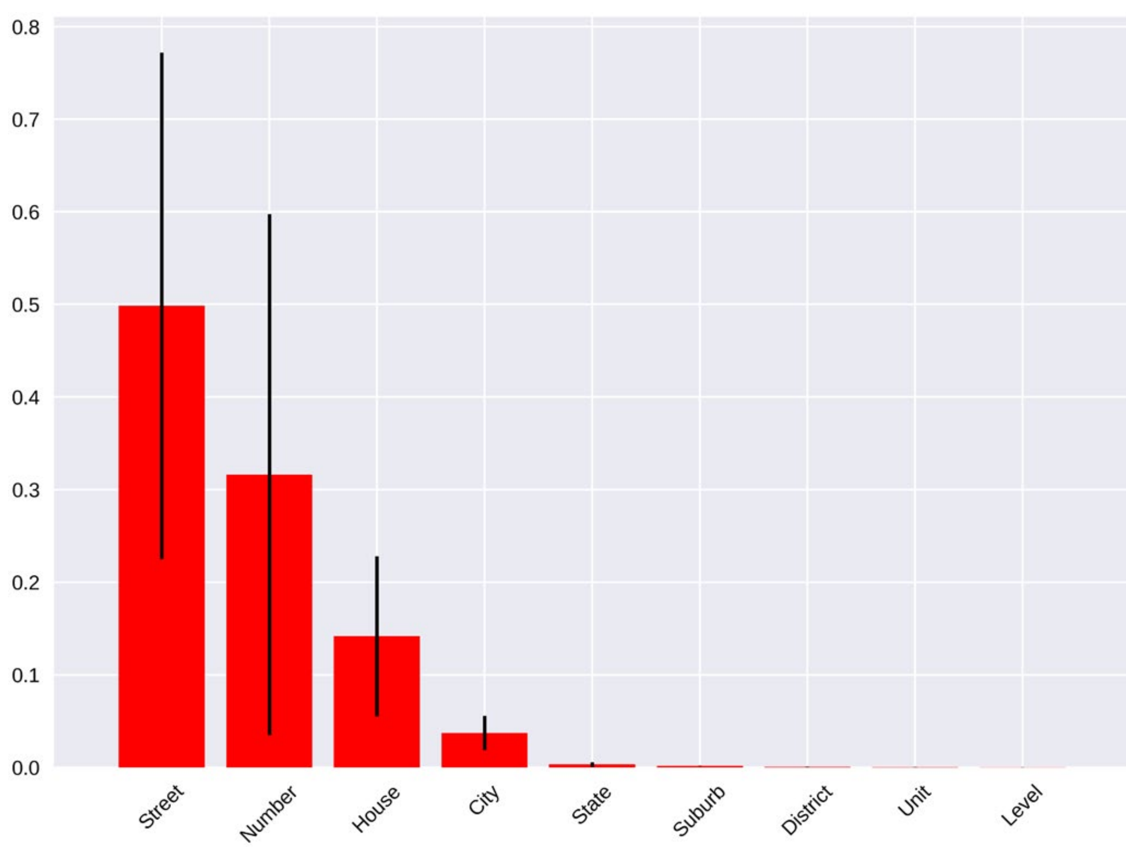


FIGURE 3 Feature importances of address fields from Table 1 to matching outcomes. Importances are given for the random forest model trained on address fields segmented by the CRFs

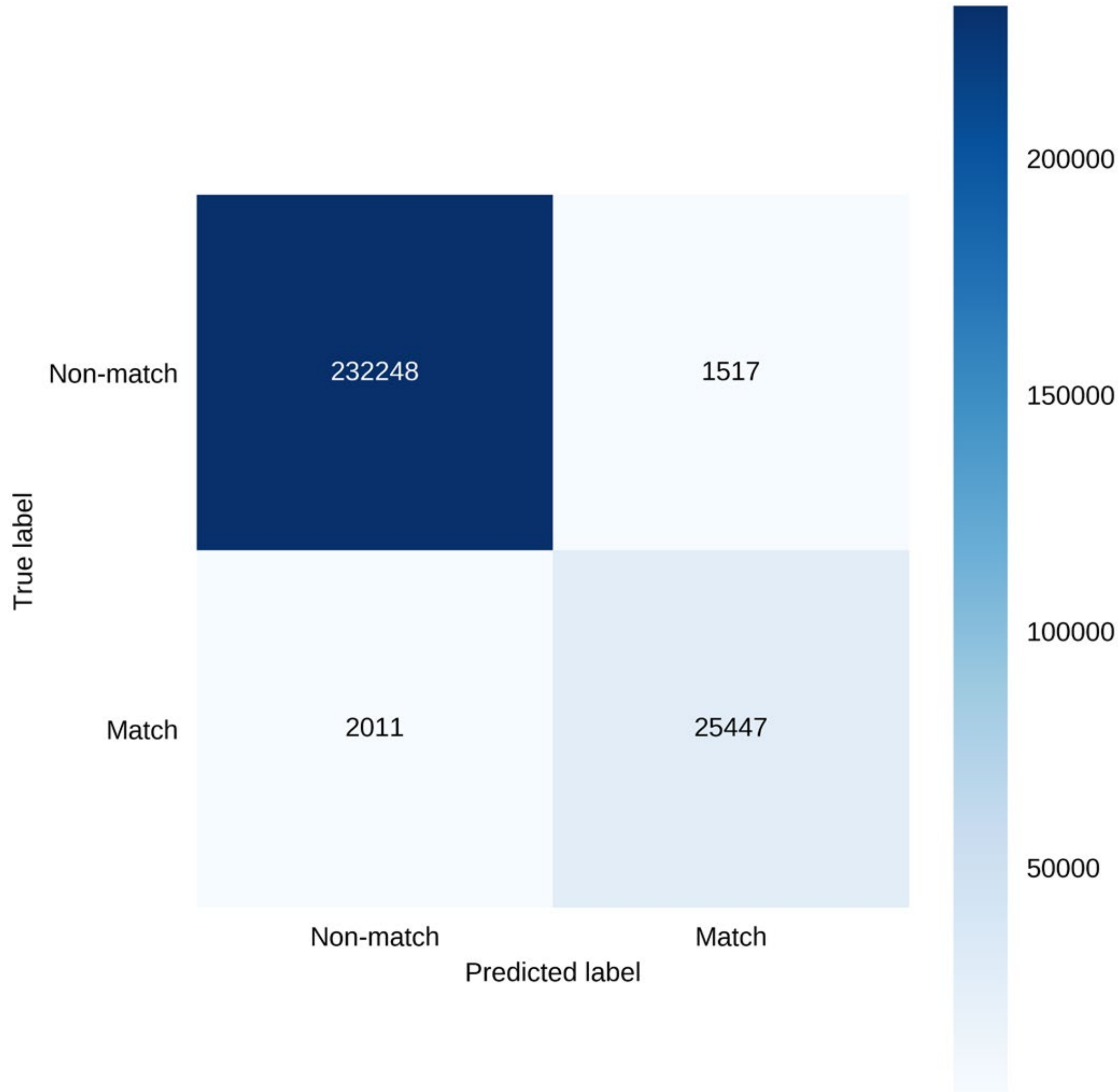


FIGURE 4 Confusion matrix for true positives, false positives, true negatives and false negatives retrieved by the random forest classifier for CRFs segmentation

value for the random forest, which we complement by displaying a precision–recall curve in Figure 5. The green line in the figure suggests that while the random forest performs well at classifying true matches from the matches it returns, it performs less well at retrieving all matched instances. Between the two ensemble approaches, XGBoost classifies the highest number of true matches correctly, with only marginal differences in recall, or retrieval of relevant address pairs, between the two. Presumably, the ensembles perform better because the vector space for classifying address pairs is not linearly separable, and requires a nonlinear decision boundary to partition matches from non-matches to a high degree of accuracy.

Next, we turn our attention to the CRFs method that is augmented with the use of word2vec for address field comparisons. While the first method uses Jaro–Winkler similarity to assess string distance between address fields parsed by the CRFs model, our second approach augments the first by replacing Jaro–Winkler similarity with cosine similarity between word vectors learnt from the parsed address fields. Overall, this augmented approach is highly competitive with the first approach. If, for example, we take the XGBoost findings from Table 3 as a point of comparison, the precision and recall values decrease by 0.005 and 0.032, respectively, in the augmented approach when compared to the first approach. This raises the question of what advantage CRFs augmentation with word-2vec yields. The principal advantage is that it does not commit the user to the biases of a particular string distance

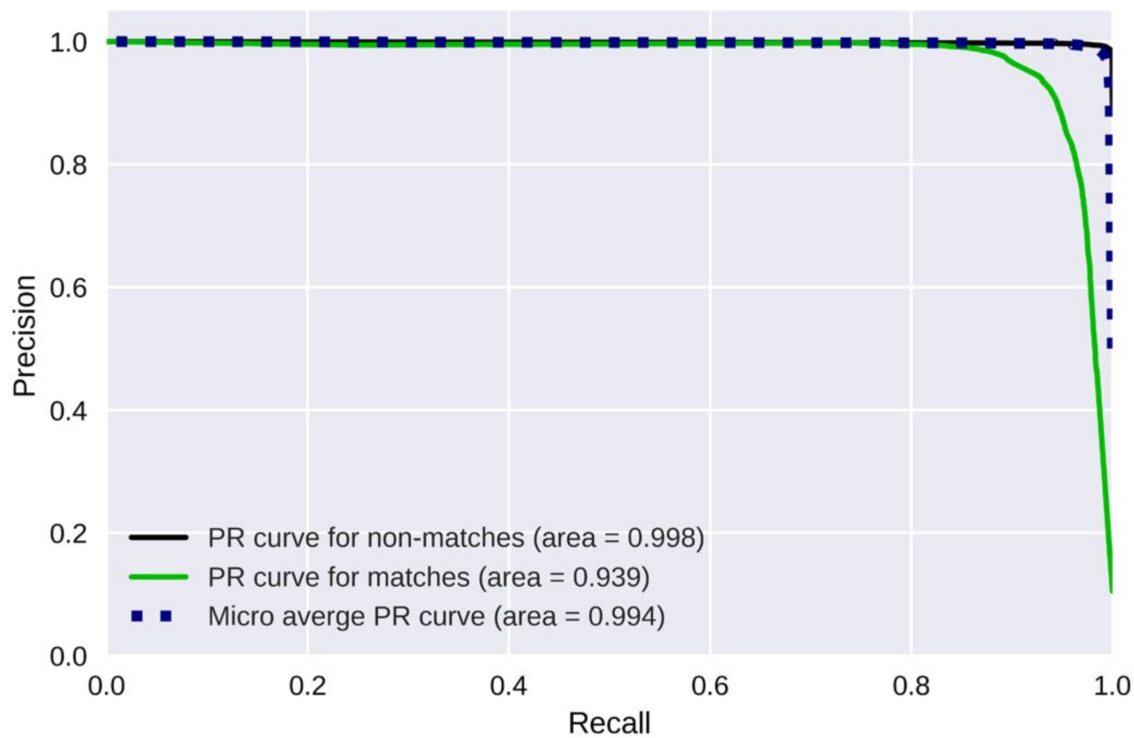


FIGURE 5 Precision–recall (PR) curve for the random forest classifier trained on addresses segmented by the CRFs model

function. Jaro–Winkler similarity, for example, is highly tuned to named attribute values, and so the use of word vectors provides a more generalizable approach for matching address fields. In all, the findings support our theoretical motivations from Section 3, and imply that the word vectors obtained for the address fields perform well in capturing the syntactic structures and regularities required to resolve address pairs to a match.

Finally, we tweak two components of our empirical design to evaluate the robustness of our main findings. In doing so, we evaluate any substantive change in the match performance for our preferred specification, the XGBoost classifier, trained on comparison vectors generated by the CRFs model. Firstly, we adjust the probabilities for introducing modifications to the synthetic non-matched addresses generated by FEBRL. Here, we apply two scenarios: in one, we set the probability of a missing field equal to the proportion of missing fields in the matched addresses, while setting the maximum modification per field and maximum modification per address to one; and in the other, we maintain the same probability for missing fields, but increase the maximum number of modifications per field and maximum modifications per address to nine in both cases. Modifications refer to insertions, deletions and transpositions of characters in the address field. These scenarios are introduced to evaluate the extent to which the classifier performance changes as we degrade the quality of our non-matches in the training data. As expected, the precision value increases marginally from 0.955 to 0.973 as we increase the number of error modifications to the non-matched addresses. This is intuitive as increasing error in the non-matches means the address pairs become more dissimilar. Therefore, it becomes easier for the classifier to disambiguate between matches and non-matches as the nuances between non-matched address pairs become less pronounced. Our second tweak tested the change of blocking mechanism from standard blocking to sorted neighborhood blocking and Q-gram blocking. This was applied to better handle cases of misspelt zip codes, which is problematic because misspellings could allocate addresses to incorrect zip code blocks. Despite the empirical motivations to alternate the blocking mechanism, we found the results of our main findings were invariant to which blocking technique was applied. Presumably this was because zip codes of the LDC and VOA addresses are of high quality, as the

VOA addresses originate from an official UK government source and the LDC have a business case in maintaining accurate, high-quality zip codes.

5 | CONCLUSIONS

Often the biggest problem when faced with spatial data is accessing it. Address matching resolves text-based address sequences to matches, integrating disparate sources of data that would otherwise remain in isolation. In this article we evaluated the performance of two recent machine learning techniques for linking address pairs where the match status was already known. Our first technique, the CRFs approach, focused on segmenting whole postal addresses into address fields, which became the basis for constructing comparison vectors for every candidate address pair in the data set. Once obtained, supervised classifiers were applied to partition the comparison vectors of address pairs into matches and non-matches. In all, the classifiers trained using addresses segmented by the CRFs achieved a precision of up to 0.955, with the ensemble learners outperforming the logistic model. This was likely due to the improved fit of a nonlinear decision boundary to the underlying vector space.

Our second approach augmented the first by replacing the string similarity metric we used to compare parsed address fields from the CRFs model with a comparison between word vectors. These were generated using a technique called word2vec, which sought to embed semantically and syntactically similar address fields to nearby locations in the vector space, with the expectation that vectors embedded *nearer* together could be used to match address fields. As before, we used supervised classifiers to facilitate the linkage, which resulted in a precision of up to 0.950. This value implied the vectors obtained for the address fields performed successfully at encoding word relationships, patterns and regularities that are required to facilitate accurate linkage between address pairs. In synthesis, the main implications of this article point to the utility of CRFs, and their augmentation with word2vec, for the accurate segmentation of addresses. These steps are preconditions for constructing high-quality comparison vectors that can be used to accurately classify address pairs into matches and non-matches.

NOTES

[1]Environmental health studies, for example, rely on spatial record linkage to determine whether individuals in residential locations live within defined zones of exposure to hazardous environments (Baldovin et al., 2015; Cayo & Talbot, 2003; Reynolds et al., 2003).

[2]Our HMM is trained on the same OSM addresses for the UK as Libpostal. They are obtained by filtering the "great-britain-latest.osm.pbf" file available from Geofabrik (2018). Filtering is performed using the *Osmosis* command line application (OpenStreetMap, 2018) that allows us to distill addresses from the entirety of OSM data in the file. Therefore, we use Osmosis to filter by the following tags: "addr:housename", "addr:housenumber", "addr:street" and "addr:postcode". Implementation for the HMM model is provided by a script available from FEBRL (Christen & Churches, 2005) which tags free-text address sequences from lookup tables before rearranging the tags to the most likely sequence of address fields.

[3]For details of technical implementation, the reader is referred to Mikolov et al. (2013).

REFERENCES

Baldovin, T., Zangrando, D., Casale, P., Ferrarese, F., Bertoncello, C., Buja, A., ... Baldo, V. (2015). Geocoding health data with geographic information systems: A pilot study in northeast Italy for developing a standardized data-acquiring format. *Journal of Preventive Medicine & Hygiene*, *56*, 88–94.

Barrentine, A. (2018). *Libpostal*. Retrieved from https://github.com/openvenues/libpostal

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, *37*, 1554–1563.

Blanchette, C., DeKoven, M., De, A., & Roberts, M. (2013). Probabilistic data linkage: A case study of comparative effectiveness in COPD. *Drugs in Context*, *2013*, 212258.

Boulos, M. N. K. (2004). Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom. *International Journal of Health Geographics*, *3*, 1–50.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Cayo, M. R., & Talbot, T. O. (2003). Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, *2*, 1–10.

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. arXiv:1603.02754.

Christen, P. (2012). *Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*. New York, NY: Springer.

Christen, P., & Churches, T. (2005). *FEBRL, version 0.3.1*. Retrieved from http://users.cecs.anu.edu.au/ Peter.Christen/Febrl/febrl-0.3/febrldoc-0.3/manual.html

Churches, T., Christen, P., Lim, K., & Zhu, J. X. (2002). Preparation of name and address data for record linkage using hidden Markov models. *BMC Medical Informatics and Decision Making*, *2*, 1–9.

Cibella, N., & Tuoto, T. (2012). Statistical perspective on blocking methods when linking large data-sets. In A. Di Ciaccio, M. Coli, & J. M. Angulo Ibanez (Eds.), *Advanced statistical methods for the analysis of large data-sets* (pp. 81–89). Berlin, Germany: Springer.

Diesner, J., & Carley, K. M. (2008). Conditional random fields for entity extraction and ontological text coding. *Computational and Mathematical Organization Theory*, *14*, 248–262.

Du, H., Alechina, N., Jackson, M., & Hart, G. (2017). A method for matching crowd-sourced and authoritative geospatial data. *Transactions in GIS*, *21*, 406–427.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge, UK: Cambridge University Press.

Ektefa, M., Sidi, F., Ibrahim, H., Jabar, M., & Memar, S. (2011). A comparative study in classification techniques for unsupervised record linkage model. *Journal of Computer Science*, *7*, 341–347.

Geofabrik. (2018). *OpenStreetMap data for Great Britain*. Retrieved from https://download.geofabrik.de/europe/great-britain/

Goldberg, D. W. (2011). Improving geocoding match rates with spatially-varying block metrics. *Transactions in GIS*, *15*, 829–850.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random forests. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The elements of statistical learning: Data mining, inference and prediction* (2nd ed., pp. 587–604). Berlin, Germany: Springer.

Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data quality and record linkage techniques*. New York, NY: Springer.

Jaro, M. (1984). *Record linkage research and the calibration of record linkage algorithms (Statistical Research Division Report Series No. Census/SRD/RR-84/27)*. Washington, DC: U.S. Bureau of the Census.

Kopcke, H., & Rahm, E. (2010). Frameworks for entity matching: A comparison. *Data & Knowledge Engineering*, *69*, 197–210.

Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In C. E. Brodley & A. P. Danyluk (Eds.), *Proceedings of the 18th International Conference on Machine Learning* (pp. 282–289). San Francisco, CA: Morgan Kaufmann.

McInnes, B. T., & Pedersen, T. (2013). Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. *Journal of Biomedical Informatics*, *46*, 1116–1124.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv:1301.3781.

Nasseh, D., & Stausberg, J. (2016). Evaluation of a binary semi-supervised classification technique for probabilistic record linkage. *Methods of Information in Medicine*, *2*, 136–143.

Newcombe, H., & Kennedy, J. (1962). Record linkage: Making maximum use of the discriminating power of identifying information. *Communications of the ACM*, *5*, 563–566.

Oliveira, G., Bierrenbach, A., Camargo, K., Coeli, C., & Pinhero, R. (2016). Accuracy of probabilistic and deterministic record linkage: The case of tuberculosis. *Revista de Saúde Pública*, *50*, 1–9.

OpenCage. (2018). *OpenCage address formatting*. Retrieved from https://github.com/OpenCageData/address-formatting

OpenStreetMap. (2018). *Osmosis*. Retrieved from https://github.com/openstreetmap/osmosis

PAF. (2018). *Postcode Address File: UK address database*. Retrieved from https://www.royalmail.com/business/services/marketing/data-optimisation/paf

Řehuřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta, Malta: ELRA.

Reynolds, P., Behren, J. V., Gunier, R., Goldberg, D., Hertz, A., & Smith, D. (2003). Childhood cancer incidence rates and hazardous air pollutants in California: An exploratory analysis. *Environmental Health Perspectives*, *111*, 663–668.

Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*, *8*, 627–633.

Singleton, A. (2015). *Data challenges and rateable value adjustment*. Retrieved from https://www.cdrc.ac.uk/case-study/data-challenges-and-rateable-value-adjustment/

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*, 260–269.

VOA. (2017). *VOA rating list*. Retrieved from https://voaratinglists.blob.core.windows.net/html/rlidata.htm

Yancey, W. E. (2005). *Evaluating string comparator performance for record linkage* (Research Report Series, Statistics #2005-05). Washington, DC: U.S. Bureau of the Census.

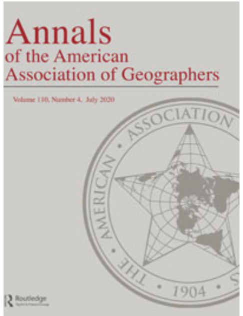
Zellig, H. (1954). Distributional structure. *Word*, *10*, 146–162.

How to cite this article: Comber S, Arribas-Bel D. Machine learning innovations in address matching: A practical comparison of word2vec and CRFs. *Transactions in GIS*. 2019;23:334–348. https://doi.org/10.1111/tgis.12522

B — Annals of the Association of American Geographers Paper

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21



Building Hierarchies of Retail Centers Using Bayesian Multilevel Models

Sam Comber, Daniel Arribas-Bel, Alex Singleton, Guanpeng Dong & Les Dolega

To cite this article: Sam Comber, Daniel Arribas-Bel, Alex Singleton, Guanpeng Dong & Les Dolega (2020) Building Hierarchies of Retail Centers Using Bayesian Multilevel Models, Annals of the American Association of Geographers, 110:4, 1150-1173, DOI: 10.1080/24694452.2019.1667219

To link to this article: https://doi.org/10.1080/24694452.2019.1667219

Published online: 22 Oct 2019.

Building Hierarchies of Retail Centers Using Bayesian Multilevel Models

Sam Comber, Daniel Arribas-Bel, Alex Singleton, Guanpeng Dong, and Les Dolega

Geographic Data Science Lab and Department of Geography and Planning, University of Liverpool

The perceived quality of urban environments is intrinsically tied to the availability of desirable leisure and retail opportunities. In this article, we explore methodological approaches for deriving indicators that estimate the willingness to pay for retail and leisure services offered by retail centers. Most often, because the quality of urban environments cannot be qualified by a natural unit, the willingness to pay for an urban environment is explored through the lens of the residential housing market. Traditional approaches control for individual characteristics of houses, meaning that the remaining variation in the price can be unpacked and related to the availability of local amenities or, equivalently, the willingness to pay. In this article, we use similar motivations but exchange housing prices for residential properties with property taxes paid by nondomestic properties to glean hierarchies of retail centers. We outline the applied methodological steps that include very recent, nontrivial contributions from the literature to estimate these hierarchies and provide clear instructions for reproducing the methodology. Using the case study of England and Wales, we undertake a series of econometric experiments to rigorously assess retail center willingness to pay (RWTP) as a test of the methods reviewed. We build intuition toward our preferred specification, a Bayesian multilevel model, that accounts for the possibility of a spatial autoregressive process. Overall, the applied methodology describes a blueprint for building hierarchies of retail spaces and addresses the limited availability of spatial data that measure the economic and social value of retail centers. *Key Words: econometrics, retail geography, spatial statistics.*

城市环境的感知质量，在本质上决定了是否可以提供理想的休闲和零售机会。在本文中，我们探索以多种方法评估在零售中心内，人们希望花钱购买所提供零售和休闲服务的意愿。最常见的情况是，由于城市环境的质量无法用简单的常规单位来衡量，因此我们需要从住宅市场的角度，研究人们为城市环境付费的意愿。传统方法涵盖了房屋的具体特征，这意味着可以使用价格中其他的变量，与当地便利设施的可用性或同等的支付意愿相关联。在本文中，我们试用了类似的动机，但替换掉了住宅地产的价格，改用非住宅项目支付的地产税，来确定零售中心的层级。我们概述了具体的方法步骤，包括最近文献中的有效性研究，对这些层级结构进行评估，为复制方法提供明确指导。利用英国和威尔士的案例研究，我们进行了一系列计量经济学实验，严格评估零售中心的支付意愿（RWTP），为了检验所提到的这些方法。我们建立了首选规范的直观模型，即贝叶斯多级模型，它解释了空间自回归过程的可能性。总体而言，所使用的方法勾勒出构建零售空间层次结构的蓝图，解决了在衡量零售中心经济和社会价值时，可用空间数据有限的问题。*关键词：计量经济学、零售地理学、空间统计学。*

La calidad que se percibe en los entornos urbanos está intrínsecamente ligada a la disponibilidad de oportunidades para el ocio preferido y de compras al detal. En este artículo, exploramos los enfoques metodológicos para derivar indicadores que calculen la inclinación a pagar por servicios de compras al menudeo y ocio que ofrecen los centros de ventas al detal. Con más frecuencia, debido a que la calidad de los entornos urbanos no puede ser calificada por una unidad natural, la voluntad de pagar por un entorno urbano se explora a través de la lente del mercado de vivienda residencial. Los enfoques tradicionales controlan las características individuales de las casas, lo cual significa que la variación restante del precio puede descargarse y relacionarse con la disponibilidad de atracciones locales o, equivalentemente, con la voluntad de pagar. En este artículo, usamos motivaciones similares, pero cambiamos los precios de las viviendas para las propiedades residenciales con impuestos a la propiedad pagados por propiedades no domésticas para deducir las jerarquías de los centros de ventas al detal. Bosquejamos los pasos metodológicos aplicados que incluyen contribuciones muy recientes y no triviales de la literatura para calcular estas jerarquías y suministrar instrucciones claras para reproducir la metodología. Usando a Inglaterra y Gales como estudio de caso, emprendemos una serie de

Annals of the American Association of Geographers, 110(4) 2020, pp. 1150–1173
Initial submission, February 2019; revised submission, June 2019; final acceptance, July 2019
Color versions of one or more figures in the article can be found online at www.tandfonline.com/raag

experimentos econométricos para estimar rigurosamente la voluntad del centro de menudeo de pagar (RWTP), como prueba del método reseñado. Construimos intuición hacia la especificación de nuestra preferencia, un modelo bayesiano de nivel múltiple que toma en cuenta la posibilidad de un proceso espacial autorregresivo. En general, la metodología aplicada describe un anteproyecto para construir jerarquías de espacios del menudeo y aboca la disponibilidad limitada de datos espaciales que midan el valor social y económico de los centros de ventas al detal. *Palabras clave: econometría, estadísticas espaciales, geografía del menudeo.*

The quality of an urban environment is a principle determinant of attractiveness (Glaeser, Kolko, and Saiz 2001). Attractiveness, in this context, might be understood as an outcome of perceived place attributes (Finn and Louviere 1996), which can be argued as those perceptions, attitudes, and patronage behavior of consumers drawn to particular places (Teller and Elms 2012). The quality of an urban environment cannot be qualified by a natural unit of analysis, however, and so approaches typically observe its capitalization into housing prices (Rappaport 2009). The depth and breadth of consumer amenities, natural and cultural assets, and opportunities in the labor market are seen as an influential driver of demand for residential space (Oner 2017). As an example, the attractiveness of Paris might be considered as a product of fine-dining restaurants, art museums such as the Louvre, and the impressive stock of buildings (Brueckner, Thisse, and Zenou 1999). Accordingly, Rappaport (2009) argued that environments with above-average consumer amenities or, implicitly, quality of urban environment typically sustain a higher density of residential population, resulting in higher prices in the housing market.

Under these assumptions, the desirability of areas has often been explored through the lens of home buyer decisions in the residential housing market. Hedonic analyses that estimate the willingness to pay for consumer amenities through residential housing markets derive a snapshot for the desirability of particular places. In recent years, the proportion of the individual's spending allocated to consuming the economy's lifestyle amenities and services has increased substantially (Oner 2017). An increasing share of the individual's rising wealth is allocated to the pursuit of enjoyment and experience, which is reflected by an increase in the willingness to pay for properties that are proximate to retail and leisure destinations. Changing consumer desires have transformed traditional retail zones into spaces of leisure consumption that are increasingly service oriented. Concentrations of retail outlets are referred to as retail agglomerations and exist across a system in space, with their attractiveness to home buyers related to the composition and richness of the retail environment but also competing opportunities available elsewhere (Teller and Elms 2012). Moreover, areas of retail perform as attractors for a multitude of heterogeneous user groups such as prospective and existing residents, consumers, visitors, and employees (Teller and Reutterer 2008). In this way, the availability of consumer amenities is seen as a driver of urban vitality, so an estimation of the willingness to pay for an amenity-rich environment can be used to gauge how desirable that area is.

One particular area that attracts a number of retail opportunities is the town center. Town centers are complex urban economic systems that are characterized by the clustering of socioeconomic activity (Thurstain-Goodwin and Unwin 2000). Embedded within the urban fabric of town centers are retail centers that are agglomerations of consumer spaces and shopping destinations that are central to economic and civic life (Pavlis, Dolega, and Singleton 2017). Town centers are typically composed of a retail center but in some cases have more expansive functional areas that include office spaces in addition to retail and services. A focus on classifying *retail center* willingness to pay (RWTP) is foundational to understanding hierarchies of retail spaces, which, by implication, reveal geographic patterns in urban growth and development. Retail center hierarchies are the rankings of particular centers within a network, the position of which relates to the size, attractiveness, and gravity of their composite retailers influence, with top-ranked centers typically offering multipurpose comparison shopping experiences that have a wider geographical reach on consumers (Dennis, Marsland, and Cockett 2002). By contrast, smaller district centers are more embedded in local economies and are patronized by a smaller catchment area. Although an underlying driver to the sustainability of the built environment, since the 1970s retail centers have become threatened by the decentralization and dispersal of development to out-of-town locations on the periphery of towns. In addition, Singleton et al. (2016) claimed that retail has become increasingly vulnerable to the effect of growing online shopping and so must be considered within a framework of e-resilience.

In this article, we introduce a statistical technique to derive indicators that describe hierarchies of retail centers across the national extent, which we obtain alongside a measure of uncertainty in the rank-ordered estimate for each retail center. Despite the concerns previously raised, although retail centers in the United Kingdom have long been examined under a series of milestone reviews (Department of the Environment Urban and Economic Development Group 1994), there is little quantitative evidence that explores the performance of town center retail economies, which has undermined effective policy formulation and decision making (Astbury and Thurstain-Goodwin 2014). Indicators of retail hierarchies produced by commercial organizations (Javelin Group 2017; CACI 2018), for example, lack fine spatial granularity at the retail center scale. Our approach is motivated by a hedonic framework of analysis that is typically oriented toward residential housing markets, except that we exchange residential for commercial properties to execute our empirical strategy. We describe the methodological steps required to reproduce the RWTP estimates, which includes very recent, nontrivial contributions from the econometrics literature. Finally, we introduce a validation exercise to verify the RWTP estimates correspond to conventional wisdom by correlating the scores to socioeconomic characteristics of the retail center. Not only is the approach we operationalize novel in application but we note that our methodology is replicable and generalizable to international contexts, conditional on data availability.

The remainder of the article is organized as follows. The next section motivates the underlying conceptual framework of the article, followed by an introduction to the specification and underlying assumptions of the modeling approach. After elaborating on the nature and limitations of the data source, we step through the results of each model, including a validation exercise to confirm whether the RWTP estimate for each retail center responds to characteristics that are associated with attractive places. The final section summarizes the article, presenting extensions for future elaborations of the applied methodology.

Background and Motivation

Modern Consumption Patterns

The desirability of urban places to live is increasingly dependent on their ability to provide consumption opportunities, which are often reflected in housing prices (Glaeser, Kolko, and Saiz 2001). Leisure and retail amenities such as restaurants, live performance venues, and shopping districts have been shown to be crucial for attracting modern workers who balance economic and lifestyle opportunity in selecting places to live and work (Florida 2000). Because perceptual qualifications for the quality of leisure and retail environments cannot be directly counted or observed, they have often been evaluated by the willingness to pay for residential property through hedonic approaches (Rivera-Batiz 1988; Hui and Liang 2016). Jin and Sternquist (2004) argued that the desire for leisure and shopping is increasingly linked to the concept of enjoyment and experience. From a consumer perspective, shopping trips not only satisfy the individual's bundle of wants and needs at a given store but they allow the consumer to speak his or her own geographies of everyday life through the language of consumption (Sack 1988). This "credit-card citizenship" toward identities and preferred lifestyle choices provides an opportunity for social mixing and participatory entertainment (Goss 1993). Over the last few years, however, this traditional brick-and-mortar retailer landscape has been restructured by the growth of electronic retailing, with e-commerce sales in the United States rising by 101 percent in the period between 2011 and 2016 (Helm, Kim, and Silvia 2018). Due to the rise of the Internet, online consumption has tilted power from retailers to consumers through opportunities for 24/7 convenience and price comparison, increased ease of market entry and transparency, and a distribution of products to a wider geographical reach (Williams 2009). Evidence suggests that this rapid expansion in online consumption has affected the health of retail centers in complex ways and has been a principal driver of change to the geography of traditional UK high streets (Wrigley and Lambiri 2014).

Adjustments as a result of online shopping to the market share of retailing, leisure, and services in retail centers are typically considered detrimental effects that cause physical shopping opportunity to be substituted online (Doherty and Ellis-Chadwick

2010). Yet, online retailing has also been linked to complementarity and modification processes that blend traditional retail channels with e-commerce by refashioning the in-store consumer experience (Poushneh and Vasquez-Parraga 2017). In the United Kingdom, major retailers including Argos, John Lewis, and Boots have integrated new technologies by opening "click and collect" points that act as points of delivery for Internet sales by allowing customers to order goods online and collect them in store (Singleton et al. 2016). Thus, the role of retail centers remains vital to modern consumption and the continuity of physical shopping environments, with consumers pointing to the hedonic experience that physical stores offer through recounted social experiences, the opportunity to discover new and exciting goods, and the gratification afforded by touching or trying products in store (Cho and Workman 2011). Under this lens, Singleton et al. (2016) recast the propensity of localized populations to engage with the mixture of online shopping and physical retailing provision under a frame work of "e-resilience." The constraint or opportunity of e-commerce to retail centers is not uniform across all retail types, with retailers whose merchandise can be replicated and digitized online the most vulnerable to large-scale store closures and lost physical shopping opportunity (Zentner, Smith, and Kaya 2013).

Geographic Behavioral Drivers of Retail Center Hierarchies

More concretely, the geodemographic characteristics of catchments served by retail centers are fundamental drivers of consumer choices and behaviors that shape the willingness to pay for retail opportunity and, in turn, hierarchies of retail spaces (Birkin, Clarke, and Clarke 2002). In the United Kingdom, geographic variation of consumer disposable incomes affects the relative retail value of catchment areas. For example, hierarchies of retail centers for large conurbations and metropolitan centers are moderated by their propensity to attract highly mobile consumers who require multiple retail and leisure choices (Wrigley et al. 2015). More generally, steps in the hierarchy of retail centers have become contingent on a rising "convenience culture." This incorporates the progressive rise of online retail with preferences for "local" shopping (and derived product authenticity, traceability, and sustainability benefits) alongside a revaluation of consumer awareness toward "community-sustaining" consumption (Chalmers et al. 2012). Since the early 2000s, significant demographic and societal shifts have driven these trends, with particular growth among low-density households, aging populations, and younger workers who are faced with longer working hours and busy lifestyles (Wrigley et al. 2015). These groups in particular have an increasing desire for convenience at the local level. In the United Kingdom this is revealed by evidence from the Institute of Grocery Distribution (IGD) that suggests that consumers are increasingly shopping little and often at shops closer to home rather than shopping at larger out-of-town retail developments, a phenomenon described as "top-up shopping" (IGD 2014). Moreover, a report by the Ethical Consumers Market suggests that the number of shoppers purchasing produce from local shops increased from 15 percent to 42 percent between 2005 and 2012 (Ethical Consumer Research Association 2013). This has considerable beneficial implications for the configuration of UK retail centers, because high streets and town centers are now increasingly the preferred locations for consumers to undertake their top-up shopping. Not only has this driven footfall back to retail centers but local shopping has reshaped hierarchies of retail spaces by boosting the vitality and viability of town centers and high streets in the United Kingdom (Wrigley et al. 2015).

Yet there is significant demographic variation in the propensity for consumers to value local shopping and engage with Internet retail; this has determined the differential geographies of online shopping (Longley and Singleton 2009) and, in turn, been an influential driver of retail hierarchies. Although typically younger age groups have been the most receptive to online shopping, significant growth has been recorded in the rate of online purchasing among those sixty-five and older, with 48 percent buying online in 2014, increasing from 16 percent in 2008 (Office for National Statistics 2018). By exploiting opportunities provided by digital technologies and adapting retail spaces to meet the needs of every population group, retail centers have become virtual marketplaces. Here, consumers are able to access information online regarding the availability of products, stores, services, and brands prior to visiting, which has enhanced the retail center customer experience (Wrigley et al. 2015). Despite these significant structural changes, though, good product ranges, quality of retail provision, and traditional factors such as overall retail center experience, atmosphere, and leisure provisions remain foundational drivers of footfall in retail centers. This extends their use from shopping destinations to areas for economic and educational activities (in addition to social interaction; Warnaby et al. 2002).

In addition to demographic variation, consumer behavioral patterns vary spatially and are directly linked to the geographies of demand toward retail facilities. Steps in the hierarchy of retail centers are intertwined with the underlying characteristics of the catchment area itself. Variations in consumer confidence, the ownership of basic digital skills, and local supply factors such as convenience and accessibility at the small-area level are influential factors toward the vitality of retail centers (Wrigley and Dolega 2011). Thus, the propensity and desirability of consumers to engage with physical shopping opportunity are governed by a multitude of contexts and influences such as the rurality and remoteness of an area (Warren 2007), the extent of Internet connectivity and speed of connection (Singleton et al. 2016), and even how informed (and educated) consumers are to access online retail (Helsper and Eynon 2010). Despite these factors, and even in a digitally transformed retail landscape, the demand for high street shops remains a permanent fixture of consumer desires, so an estimation of the willingness to pay for retail centers is foundational to unpacking hierarchies of retail spaces that reveal geographic patterns in urban growth and development.

Measuring Attractiveness

Within the academic literature, measures for estimating attractiveness[1] are most typically classified into two streams of research. Models of the first stream are inspired by Reilly's (1931) gravitational law of retail, which motivated the seminal work of Huff (1963). The Huff model applies Newtonian laws of physics to estimate a retail catchment area that factors in the spatial distribution of competing retail destinations when evaluating their gravity or consumer pull to different population groups (Dolega, Pavlis, and Singleton 2016). Huff models are advantageous because they simultaneously estimate break points in the demand surface for all competing retail destinations in the model and reduce the probability of a consumer to patronize a given location to three groups of variables, namely, distance between shops and consumers' homes; a measure of attractiveness such as store size, service levels, or opening hours; and competition proxied by the number of retail units in a location (Teller and Reutterer 2008). Yet, the usual criteria for retail attraction in Huff models are often argued as incomplete, because additional factors that affect the consumer's propensity to visit a retail destination involve a suite of qualitative indicators such as the variety of retail tenants; site-related factors such as accessibility and parking conditions; and environmental factors reflected by sensual stimuli such as ambience, atmosphere, and perception of safety (Teller and Elms 2010). Clearly, these indicators influence the choice of shopping destination, but measuring across a national extent is difficult (Dolega, Pavlis, and Singleton 2016).

Methods of the second stream are motivated by findings that demonstrate that housing prices increase faster than wage levels, implying a premium for particular locations (Glaeser, Kolko, and Saiz 2001). This has led to a number of studies estimating the relevance of consumption opportunities to the desirability of places, with a focus on home buyer decisions toward urban amenities. That is, by controlling for property-specific characteristics of a residential property such as the number of bedrooms or bathrooms or whether the property has a garden, the residual variation in the property value can be unpacked and related to the local availability of amenities or lifestyle opportunity. Using this approach, the desirability of urban environments has been shown to be factored into property values and is broadly defined by the provision of place-specific assets and amenities that contribute to the allure of an urban area (Brueckner, Thisse, and Zenou 1999). Its importance, therefore, is intrinsically tied to population growth and development (Glaeser, Kolko, and Saiz 2001; Clark 2003), because attractive places that elevate one's experience of an urban environment through concentrations of arts, leisure, and retail have been shown to attract highly skilled individuals (Florida 2008). Clark (2003), for example, demonstrated that university graduates are more likely to locate in areas with high numbers of constructed amenities such as museums, libraries, and leisure outlets. Oner (2017) paid particular attention to the role of retail as an urban amenity, regressing a Q-ratio—a ratio of the marginal price of a property to the marginal production cost—on variables reflecting accessibility to shopping destinations. In all, the study found a significant increase in the Q-ratio of 0.1 for every 1 percent increase in the accessibility to shops for city municipalities.

Measuring Retail Center Attractiveness

In this article, we follow methods of the second stream. Thus, we apply a hedonic framework to estimate the willingness to pay for retail centers. Given our focus on retail environments, business rates paid by commercial property such as high street shops provide an alternative, yet more suitable, lens to explore hierarchies of retail centers than housing prices; although rent or housing prices are our idealized data set, these are difficult to obtain, particularly at the national level. With motivations similar to the way urban economists proxy willingness to pay through residential housing, by controlling for property-level characteristics in business rates—the total floor area, the number of car parking spaces, the store type, for example—the remaining variation in a premise's business rate can be explained by home buyer desirability for a particular area or, in our case, the retail center. In the United Kingdom, nondomestic rates, or business rates, are a property-based tax levied on the estimated value of all nonresidential properties such as shops, offices, warehouses, and factories (Adam and Miller 2014). Business rates are determined using a ratable value for each nondomestic property. This is set by the Valuation Office Agency (VOA), which analyzes rent evidence (rent and lease agreement details) in addition to undertaking visual inspections of properties to ensure that all evidence is considered fairly. VOA surveyors set ratable values to reflect features including total floor area; business assets such as lifts, air conditioning, and closed-circuit television (CCTV) security systems; and changes in the local property market (VOA 2014). A valuation begins by setting a common basic value per square meter for similar properties in the same area. This basic value is then adjusted to reflect the property's individual features. Each review of a property's valuation considers property-level characteristics and, most important, the buoyancy of the local property market. In this way, business rates are synchronized to local economic market conditions, reflecting the relative size and scale of retail economies (Astbury and Thurstain-Goodwin 2014).

In our study, we label the estimated phenomena as RWTP, which describes the price that home buyers ascribe to the leisure and retail services offered by retail centers proximate to the property. In all, our findings are permissible because the residual variation in the business rate is attributed to local property market conditions (VOA 2014), which themselves are influenced by home buyer aspirations to reside in an environment that satisfies their wants and desires (Glaeser, Kolko, and Saiz 2001). By implication, this means that the ratable value, once controlling for property-level characteristics, can be used to approximate RWTP for the retail center with a catchment that services the surrounding area. Using our conceptual approach, we can begin to unpack hierarchies of retail centers by undertaking a series of experiments on several econometric techniques to find a preferred specification that provides the most rigorous estimates of RWTP for retail centers across the case study of England and Wales.

Methodological Framework

Our approach to estimate RWTP relies on hedonic modeling (Rosen 1974). This technique is typically used in the real estate literature to disentangle the price of a complex good as a function of the multiple intrinsic and extrinsic characteristics common to the property. In our case, a hedonic framework is applied to unpack the determinants of business rates for individual stores. By controlling for various property-level descriptors, a hedonic approach that uses a variable to represent each retail center allows us to recover the implicit price for the retail and leisure opportunities provided by the retail center. Practically speaking, this approach translates into a regression that explains the willingness to pay for receiving consumer amenities inside different retail centers. Once controlling for property-specific characteristics, the RWTP effect can be recovered for the location where stores are located because the business rate for each property involves setting a common basic value per square meter for similar properties in the same area, reflecting the performance, size, and scale of local market conditions (Astbury and Thurstain-Goodwin 2014).

To estimate the most robust empirical hedonic model specification, we compare several approaches, with a focus on recent contributions to the literature. To begin, we introduce a baseline spatial fixed effects model (Anselin and Arribas-Bel 2013), which is expressed as

$$\ln y_{ij} = \boldsymbol{x}_{\mathbf{ij}}'\boldsymbol{\beta} + \sum\nolimits_{j=1}^{J} \theta_j D_j + \epsilon_{ij}, \tag{1}$$

where y_{ij}, the business rate for each store i in retail center j, is log-transformed to alleviate the potential impact of heteroskedasticity; $\boldsymbol{x}_{ij}'$ is a $1 \times k$ vector of store-level variables in the Appendix, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficient to be estimated; D_j is the dummy variable for retail center membership where $D_j = 1$ when $j = h$ for $i \in h$, 0 otherwise; and ϵ_{ij} is the model residual term, following an independent normal distribution $\mathcal{N}(0, \sigma_e^2)$. For model identification, the intercept is constrained to equal zero so that a separate RWTP effect θ_j can be estimated for each retail center. From a nontechnical standpoint, θ_j can be interpreted as the average willingness to pay (in log units) for stores to market their services in retail center j. One might expect different retail centers to offer varying degrees of utility such as access to particular socioeconomic groups, amount of footfall, or the prestige of surrounding consumer amenities. Taking into account individual store characteristics, θ_j captures the RWTP of retail centers.

Limitations exist associated with the fixed effect estimation strategy for the RWTP. First, the estimator of θ_j, $\hat{\theta}_j$, would not be reliable and precise if the number of stores in retail center j, n_j, is small. In addition, if different spatial processes operate at the property and retail center scale, the conflation of unobservable influences will violate the independence of errors assumption through heteroskedastic or spatially correlated error in the covariance structure (G. Dong and Wu 2016). Multilevel models are approaches that allow variance between areas, so they remedy these issues by treating the retail center as part of the explanation for geographically varying outcomes (Owen, Harris, and Jones 2016). Instead of fitting a spatial fixed effect that assumes the relationship between the predictors and response holds as constant, multilevel models factor both spatial heterogeneity (differences) between areas and also dependencies (similarities) within them (Jones 1991). Put another way, this allows two stores located within the same retail center to be more alike in their outcomes than would be expected given their individual characteristics alone. Correlation within boundaries is expected because stores are assumed to be affected by the same aggregate effects, also known as group dependence (G. Dong and Harris 2015). Our second model thus requires a two-level hierarchical structure, an outcome variable measured at the lower level geography—individual stores—and a more aggregate spatial scale for the higher level—retail centers. We specify a random intercept multilevel model as

$$\begin{gathered} \ln y_{ij} = \boldsymbol{x}_{\mathbf{ij}}'\boldsymbol{\beta} + u_j + \epsilon_{ij} \\ \mathrm{var}(\epsilon_{ij}) = \sigma_e^2; \mathrm{var}(u_j) = \sigma_u^2, \end{gathered} \tag{2}$$

where u_j ($j = 1, 2, \ldots, J$) measures the RWTP of the retail center j, assumed to be independently distributed as $\mathcal{N}(0, \sigma_u^2)$. Under Equation 2, the dependency between stores in the same retail center j is

$$cov(y_{ij}, y_{ij}) = cov(u_j + \epsilon_{ij}, u_j + \epsilon_{ij}) = \sigma_u^2. \tag{3}$$

The random intercepts u_j are a linear combination of fully pooled and no-pooling models. The fully pooled model ignores heterogeneity by fitting a common intercept for all retail center boundaries, whereas the no-pooling model, identical to the spatial fixed effect, assumes a separate intercept for each retail center. The multilevel model introduces the partial pooling, or shrinkage, of the RWTP effect toward the global intercept (Gelman and Hill 2007). This is expressed as

$$u_j = \tau_j u_j^{NP} + (1 - \tau_j) u_j^{FP}, \tag{4}$$

where u_j can be seen as a compromise between the no-pooling estimate u_j^{NP}, where each retail center is assigned its own indicator variable, and the fully pooled estimate u_j^{FP}, which assumes a single intercept for all retail centers. This precision-weighted compromise is governed by the shrinkage factor τ_j (e.g., Goldstein 2003),

$$\tau_j = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2 / n_j)}, \tag{5}$$

where the weighting for τ_j is determined by the sample size in the jth retail center (n_j) and the variation within (σ_e^2) and between (σ_u^2) groups (Goldstein 2011). For example, when a retail center's boundaries contain a small number of stores n_j, the RWTP estimate is pulled toward the fully pooled estimate. Similarly, when the boundary-level variance σ_u^2 is small—when the RWTP of retail center boundaries is similar—estimates are pooled more toward the mean level than when σ_u^2 is large.

The use of a multilevel model in the estimation routines for constructing hierarchies of retail centers represents a novel application. Multilevel models have been used to produce league tables by inferring school effectiveness from individual pupils' educational attainment but, to our knowledge, have never been applied to explore hierarchies of retail centers. Moreover, although this area of social science has a rich history in the direct application of multilevel models (Goldstein 2003), they rarely account for explicit spatial hierarchies in the empirical design. Thus, there has been emerging interest in incorporating spatial dependence into multilevel models (G. Dong and Harris 2015). Although we pursue a modeling objective similar to educational research by building a league table of retail centers, in the remainder of this section we develop an empirical strategy that accounts for potential spatial autocorrelation across the system of retail centers in space.

The model specified in Equation 2 adopts a deterministic, container-driven view of geographical space that contrasts with the reality that two retail centers located close together might be similar given their spatial proximity (Owen, Harris, and Jones 2016). In our case, we expect the RWTP effect induced by the retail center at a particular location to be directly dependent on observed values at surrounding locations, with the intensity of this influence moderated by geographic proximity. This interaction is described by a simultaneous autoregressive (SAR) process. If the data-generating process contains inherent spatial correlation, this could bias the estimated variance used for statistical inference. To account for this possibility in a spatially explicit hierarchy, G. Dong and Harris (2015) distinguished between two kinds of spatial dependence: horizontal and vertical. The *horizontal* are the spatial dependencies between lower level units that are the traditional concern of spatial econometrics (Anselin 1988), and the *vertical* are top-down group dependencies due to regional effects. One potential problem is the vertical spatial dependence effect that causes the RWTP effect in nearby retail centers to be more similar than those further away. To account for this possibility, we specify a hierarchical spatial autoregressive (HSAR) model (G. Dong and Harris 2015) that integrates SAR processes for the higher level residuals:

$$\begin{gathered} \ln y_{ij} = x_{ij}\beta_k + \theta_j + \epsilon_{ij}, \\ \theta_J = \lambda M_j \boldsymbol{\theta} + u_j, \end{gathered} \tag{6}$$

where M is a $J \times J$ spatial weights matrix that captures the interaction structure of stores by assigning nonzero weight $M_{ij} \neq 0$ to pairs of observations assumed to be spatial neighbors and zero otherwise. M_j is the jth row of M. Given the spatial characteristics of the data points, we define neighbors using an exponential decay function with the distance bandwidth d set to 5 km.[2] Following convention, M is row-standardized so that each row sums to unity $\sum M_{ij} = 1$. The parameter λ quantifies the correlation of RWTP, with higher values for λ leading to spatial covariance that dissipates slower for a higher order of neighbors. The reduced form of $\boldsymbol{\theta}$ in Equation 6 is

$$\boldsymbol{\theta} = \left(I_j - \lambda M\right)^{-1}\boldsymbol{u}, \boldsymbol{u} \sim \mathcal{N}\left(0, I_J \sigma_u^2\right),$$

where the spatial filter $(I_J - \lambda M)^{-1}$ captures any vertical spatial dependence in the RWTP effect θ_j. A Leontief expansion of the matrix inverse expands to $(I_J - \lambda M)^{-1} = I + \lambda M + \lambda^2 M^2 + \lambda^3 M^3 + \cdots$ and demonstrates spatial feedback when an increasing order of neighbors creates bands of ever larger reach around each location, relating every retail center to every other one (Anselin 2003).

A different, but related, model we specify next is a hierarchical spatial error (HSE) model, which is similar, except that we specify a spatially autocorrelated error term in η,

$$\begin{gathered} \ln y_{ij} = x_{ij}\beta_k + \theta_j + \epsilon_{ij}, \\ \theta_j = u_j + \lambda M \boldsymbol{\eta}. \end{gathered} \tag{7}$$

A final methodological consideration relates to Lesage's (2014) empirical question as to whether the spatial process under study is global or local. The covariance structure induced by the HSAR model is global, because the spatial process relates every retail center to each other one. A hierarchical spatial moving average (HSMA) process, on the other hand, considers only first- and second-order neighbors, beyond which the spatial covariance is zero (Anselin 2003):

$$\theta_j = \gamma M \boldsymbol{\theta} + u_j. \tag{8}$$

The data-generating process of Equation 8 collapses to the reduced form

$$\theta_j = (I_J + \gamma M)\boldsymbol{u}, \tag{9}$$

where everything holds as in Equation 6, except that we introduce the HSMA parameter γ. Unlike Equation 6, because $(I_J + \gamma M)$ is not inverted in the

HSMA specification, there is only local range for the induced spatial covariance. This approach is intuitive, because there might only be local interaction across a neighborhood of different retail center boundaries, as opposed to interaction across the entire system of the national extent.

Whereas the standard multilevel model is estimated using maximum likelihood estimation, the spatial models are estimated using a Markov chain Monte Carlo (MCMC) simulation technique, the stationary distribution of which constructs a target probability distribution for the parameters. MCMC simulations are typically the only feasible approach for fitting spatial models that introduce the complexities of place relatedness into the variance–covariance structure (Lesage 1997). With these motivations, conditional Gibbs samplers are derived for the HSAR (G. Dong and Harris 2015) and HSE and HSMA (Wolf et al. 2018) models to obtain posterior samples for each parameter. This way, the joint density for the parameters is broken into univariate conditional probabilities where every successive parameter draw is conditioned on the draw for the previous parameter value (Geman and Geman 1984). Not only is this sampling technique computationally efficient but the draws from the parameter space $\{\beta, \sigma_e^2, \sigma_u^2, \lambda\}$ accumulate to an entire distribution for each parameter. In our case, we summarize each parameter estimate by the median value across the distribution but also with interval calculations. Each sampling chain is simulated for 10,000 iterations, with the first 5,000 draws discarded as "burn-in" to allow the posterior distributions for each parameter to converge. In addition, we assess the serial autocorrelation for the posterior draws by examining the effective number of independent samples. As in time series analysis, we evaluate this because autocorrelation can often understate estimates of the variance in correlated sequences. On a final methodological note, the same weakly informative prior distributions were assigned to the model parameters in each model.[3] Further details on the technical implementation of the spatial models for the HSAR are found in G. Dong and Harris (2015) and for the HSE and HSMA in Wolf et al. (2018).

Despite our empirical strategy becoming increasingly sophisticated, a commonality between each model is that we obtain a free measure of uncertainty alongside each estimate of the willingness to pay for a particular retail center, θ_j. Uncertainty is expressed in the estimates for the confidence intervals of the spatial fixed effect and multilevel models and Bayesian credible intervals for the HSAR, HSE, and HSMA models. Because point estimates for θ_j represent an absolute ranking, overlapping interval estimates for each retail center imply confidence or credibility regions that change the rank-ordered estimate of centers in the hierarchy. Where the density bands of the confidence or credible intervals become less disjoint, there is increased uncertainty in the disambiguation between ranks of a given set of retail centers. Uncertainty measurements are desirable in cases where retail centers contain a small number of stores n_j. Returning to Equation 4, because this carries implications for the calculated u_j, an uncertainty estimation is valuable to ascertaining a measure of trust in the rankings of retail centers.

Data

Our point of departure for the proposed methodological approach is a geographical data set sorted into a hierarchical structure consisting of units grouped at two different levels. The points in our lower level geography represent 355,076 individual high street stores across England and Wales that are located inside retail center boundaries. This includes franchised chains such as fast-food outlets, supermarkets, and clothing stores—McDonald's, Tesco, and Primark, for example—but also independent retailers with more local scope. These data were collected by a large pool of surveying teams from the Local Data Company (LDC) in 2015 and include various descriptors for each property such as retail function and occupancy status. The most important characteristic of the data is that commercial addresses in the LDC database are matched to addresses in the VOA 2010 rating list (VOA 2018). This affords us a business rate valuation for every nondomestic premise, allowing us to unlock a rich, unique, and highly granular data set that provides a new and alternative lens through which to explore the implicit value describing the willingness to pay of an area. In all, for every store we have store-level variables that offer a rich description of the premise's physical condition. This includes data collected by VOA surveyors on the date of assessment such as the total floor area, the number of rooms in the premise, and the number of car parking spaces but also data collected by the LDC that categorize the business's function.[4] A full description of the variables that enter our design matrix is provided in the Appendix.

There are limitations to the VOA rating list that introduce error, though, especially given the primary purpose of the list is not intended for data analysis. The most notable limitation is what Astbury and Thurstain-Goodwin (2014) described as the regional difference in data collection techniques that affect the extent to which the ratable value reflects the market tone of a particular area and could lead to over- and underpredictions of the business rate assessed for the premise. Moreover, although the rating list was released in 2010, the ratable values set are actually conditioned on the 2008 market climate. Given that the UK economy underwent the shock of an economic crisis during this period, a time characterized by fragile consumer confidence, a decline in household disposable incomes, and rising shop vacancy on the high street (Department of Business, Innovation and Skills [BIS] 2011), it is likely that the overall market tone has been over- and undervalued across retail centers for England and Wales. Despite these considerations, the VOA ratings list provides highly granular and geographically accurate access to data reflecting local market economic conditions for the national extent.

The retail center is the observational unit from which we obtain home buyer willingness to pay estimates. Our higher level units are represented by 2,951 exogenously determined retail centers across England and Wales. Conceptually, retail centers are an appropriate choice for this purpose because they are drivers of local economic performance and reflect the wider economic health and social well-being of the urban environment (BIS 2011). Moreover, although they are often viewed as hubs for retail activity, they also exhibit a multitude of heterogeneous uses, including services, offices, and residential and public buildings (Teller and Elms 2012). The boundaries used in this study were produced by Pavlis, Dolega, and Singleton (2017) as a successor to boundaries developed by Thurstain-Goodwin and Unwin (2000) for the Department for Communities and Local Government (DCLG) in 2004, with the exception that they were intended to move away from a definition of town center locations of employment to functional spaces delineated for retail. Although the resulting retail centers might not perfectly align with those designated in governmental planning policy, they provide a consistent method for comparing retail centers nationally. In all, these boundaries are our higher level geographical unit and represent the functional economic market area of the retail center. The resulting spatial hierarchical structure of the data is illustrated in Figure 1 through the example of Liverpool.

Empirical Findings

In this section, we develop a discussion of our empirical findings in two main directions: First, we step through each of the modeling approaches, building intuition toward our preferred specification; second, we introduce a validation exercise to evaluate whether variation in the estimated RWTP effect can be attributed to characteristics that are generally associated with attractive areas.

Model Validation

Our point of departure is a discussion of the results provided for in the proposed methodology.[5] Thus, before exploring the subtleties of the multilevel specifications, we first step through a description of the parameter estimates for the store-level explanatory characteristics. To do this, we use the classical multilevel model (shown by the second column in Table 1) as a baseline but note that the estimates are generally consistent across each model. Overall, the estimates for the store-level covariates that enter our design matrix are fairly intuitive and of the expected signs for all models. For example, every additional room in the premise increases the ratable value by 7 percent, which is consistent with the VOA's mandate to adjust the ratable value by property-level characteristics (VOA 2014). This is also reflected in the number of car parking spaces, where each additional ten spaces increases the ratable value by 1.3 percent. Somewhat surprising, increasing the total floor area by 1,000 m^2 only seemed to increase the ratable value by 2.3 percent, but given that we control for different store sizes latently with the store category variables, this is somewhat expected. On the whole, the store type categorizations are consistent with conventional wisdom. The ratable value for premises such as takeaway food outlets, for example, is generally 20.2 percent less than the reference category, showrooms. This makes sense because the locations of takeaway outlets are generally linked to

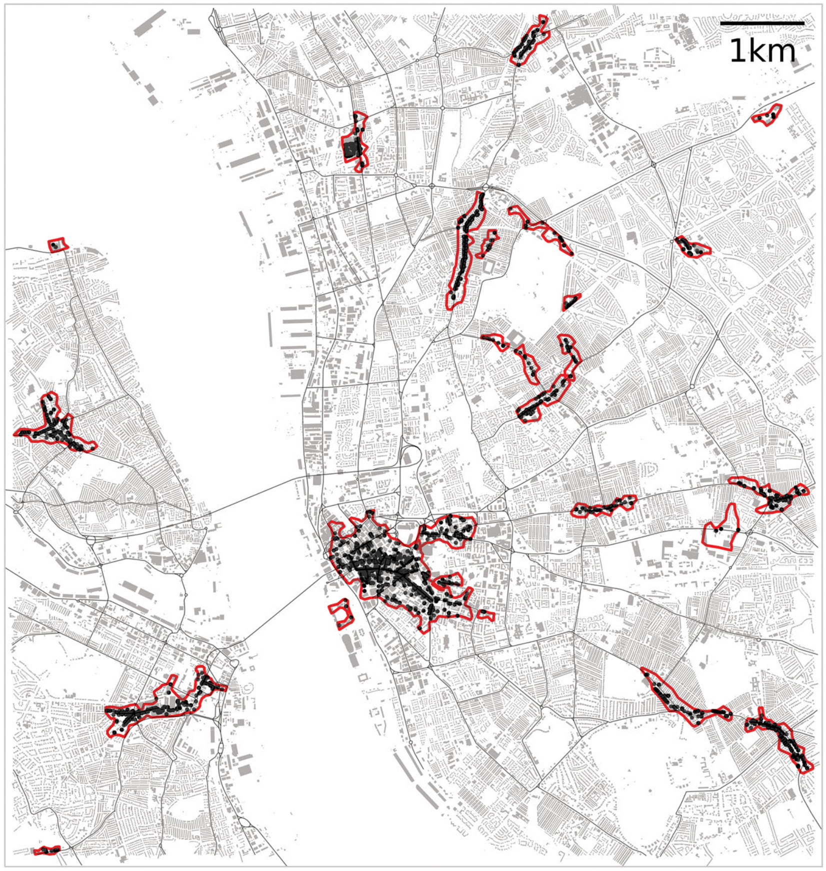


Figure 1. Lower level store premises nested into higher level retail center boundaries for Merseyside, UK.

geographical inequalities in health outcomes (Daras et al. 2018), which are simultaneously related to environments that are considered less desirable. On the other end, the ratable value for hypermarket stores (with a gross floor area over 2,500 m^2) is three times greater, which is expected given the number of business assets such as lifts, warehouse machinery, and CCTV security systems common to large supermarket stores.

We next address model selection by means of goodness-of-fit tests. In each case, every model had a highly similar root mean square error (RMSE) and log-likelihood value. Although the R^2 of the spatial fixed effect model (67.6 percent) is marginally higher than the multilevel model(s) (67.1 percent to 67.2 percent), the spatial fixed effect fits a parameter for $J = 2,951$ retail centers, which contrasts with the regularization introduced by hierarchical pooled effects in the multilevel models for smaller groups. In other words, not only does the spatial fixed effect likely overfit but the estimates and standard errors of the retail center fixed effect will be noisier in places with a smaller number of properties. In our case this is pertinent because the minimum number of stores across retail center boundaries is two. For this reason, we motivate our preferred specification as the multilevel model(s). Because the performance of each multilevel specification is comparable on goodness-of-fit grounds, however, we undertake further examination of the substantive effects in the RWTP estimate later on.

A comparison of the rank-ordered estimates for the RWTP effect θ_j are visualized for each model in Figure 2, which reflects our rankings of point estimates for RWTP, along with a measure of uncertainty shown by the 95 percent confidence (fixed effect and multilevel model) and credible (HSAR, HSE, HSMA) intervals. If any of the confidence or credible density bands for any two models overlap, the two estimated ranks are not distinct. The rankings, 1 to 2,951, are presented on the x-axis, and the y-axis displays the estimated RWTP value in log units. Additionally, we include a zoomed inset to highlight movement in the estimated RWTP value,

Table 1. Regression coefficients estimates for estimated models

	Dependent variable Ln business rate				
	FE (1)	MLM (2)	HSAR (3)	HSE (4)	HSMA (5)
(Intercept)	N/A	9.088***	9.064***	9.067***	9.071***
	N/A	(0.015)	(0.018)	(0.026)	(0.027)
Structural characteristics					
No. rooms	0.070***	0.069***	0.069***	0.069***	0.069***
	(0.000)	(0.0003)	(0.0003)	(0.000)	(0.0003)
Floor area	0.023***	0.023***	0.023***	0.023***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Car parking spaces	0.013***	0.013***	0.013***	0.013***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Store typology					
Banks and other A2 uses	0.048***	0.046***	0.048***	0.046***	0.046***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Factory shops	0.060***	0.057***	0.058***	0.057***	0.057***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Food stores	0.066***	0.061***	0.061***	0.061***	0.061***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Hairdressing/beauty salon	−0.356***	−0.360***	−0.360***	−0.360***	−0.360***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Hypermarket/superstore	3.037***	3.053***	3.054***	3.053***	3.054***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Large food stores	1.556***	1.568***	1.569***	1.569***	1.569***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
Large retail shops	2.770***	2.770***	2.766***	2.764***	2.760***
	(0.427)	(0.427)	(0.425)	(0.429)	(0.427)
Nonretail	−0.385***	−0.388***	−0.388***	−0.388***	−0.388***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Other	−0.312***	−0.316***	−0.316***	−0.316***	−0.316***
	(0.0125)	(0.012)	(0.013)	(0.013)	(0.013)
Pharmacies	−0.197***	−0.200***	−0.200***	−0.200***	−0.200***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Post offices	0.001	−0.003	−0.002	−0.003	−0.002
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Restaurants and bars	−0.051***	−0.051***	−0.051***	−0.051***	−0.051***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
Retail shops	0.108***	0.106***	0.106***	0.106***	0.106***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Takeaway food outlet	−0.199***	−0.202***	−0.202***	−0.202***	−0.202***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Variance components					
σ_e^2	0.541	0.541	0.541	0.541	0.541
	(0.0001)	(0.0001)	(0.001)	(0.001)	(0.001)
σ_u^2	N/A	0.503	0.484	0.477	0.492
	N/A	(0.0001)	(0.014)	(0.014)	(0.014)
λ	N/A	N/A	0.232***	0.230***	0.189***
	N/A	N/A	(0.026)	(0.023)	(0.020)
RMSE	0.732	0.733	0.747	0.733	0.732
Pseudo-R^2	0.676	0.671	0.671	0.671	0.671
Log-likelihood	395,363.5	402,548.3	395,478.8	395,481.9	395,479.2
Akaike's information criterion	796,644.9	805,136.7	790,993.7	790,999.7	790,994.4

Notes: p Values for Bayesian models correspond to credibility intervals crossing zero. FE = fixed effect; MLM = multilevel model; HSAR = hierarchical spatial autoregressive model; HSE = hierarchical spatial error model; HSMA = hierarchical spatial moving average model; RMSE = root mean square error.

*$p < 0.1$.

**$p < 0.05$.

***$p < 0.01$.

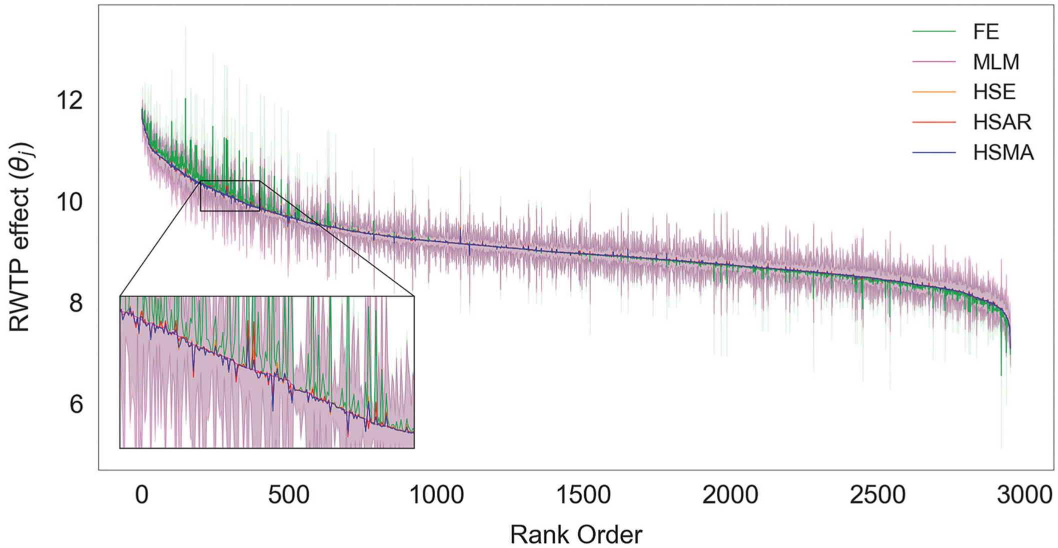


Figure 2. Rank-ordered estimates for retail RWTP. *Note:* Ninety-five percent confidence and credible intervals are shaded either side of the point estimate. FE = fixed effect; MLM = multilevel model; HSE = hierarchical spatial error model; HSAR = hierarchical spatial autoregressive model; HSMA = hierarchical spatial moving average model; RWTP = retail center willingness to pay.

which is zoomed at a window that displays the most variability in the estimated scores between each model. Taking a closer look, it appears that the movement for the RWTP estimate relative to the spatial fixed effect model, marked by the green line, is not uniform. In the upper and lower tails, for example, there is systematic variation in the parameter estimates between the spatial fixed effect and estimates of the multilevel models. This suggests that the point estimates for RWTP values deviate widely from the multilevel models for the most and least desirable retail center boundaries, with little systematic variation in between. At a general level, Figure 2 reproduces a classical result, because the estimates of the multilevel model demonstrate hierarchical pooled effects; that is, shrinkage toward the global intercept. Here, the estimates exhibit improved precision, which contrasts with the higher magnitude of uncertainty in the spatial fixed effect estimates, as shown by the more extreme and noisy estimates in the upper and lower tails of the figure. Shrinkage effects can be seen clearer in Figure 3, where we sample nine retail centers from our rankings to demonstrate movement in the RWTP estimates by expanding the point estimates horizontally along a two-dimensional axis for each model. In the case of Meridian Leisure Park, Leicester, for example, the fixed effect estimate is shrunk from 10.42 ± 0.73 to 9.72 ± 0.51 in the MLM. In real terms, this reflects a change in magnitude from £33,523.43 to £16,647.24 when we exponentiate from log units. Interestingly, what is also observable for this retail center is what Wolf et al. (2018) described as "spatially-local shrinkage," where spillovers from the *j*th adjacent retail centers cause growth in the spatial multilevel estimates toward the mean of neighboring retail centers from 9.72 ± 0.51 to 9.82 ± 0.51 under the HSAR model. Although none of the interval estimates become disjoint for each retail center, the findings from the spatial models suggest that the RWTP estimate is moderated by shrinkage toward the values of neighboring retail centers.

Having discussed our rankings, we now build intuition toward our preferred specification for the RWTP estimate, which we begin by turning our attention to the within-boundary (σ_e^2) and between-boundary (σ_u^2) variance components. By combining these measures, we calculate the variance partitioning coefficient (VPC) for the multilevel model, which measures the proportion of variance explained by the hierarchical structure ($\frac{\sigma_u^2}{\sigma_u^2+\sigma_e}$). This measure outlines the correlation between stores within the same retail center and is required to ascertain the percentage of variation explained by the retail center differences for store *i* in retail center *j* (Browne et al. 2005). The VPC statistic reveals a value of 0.482, meaning that 48.2 percent of the variance in the response is explained by the retail center geography. This VPC value motivates the empirical decision to take our multilevel models as the preferred

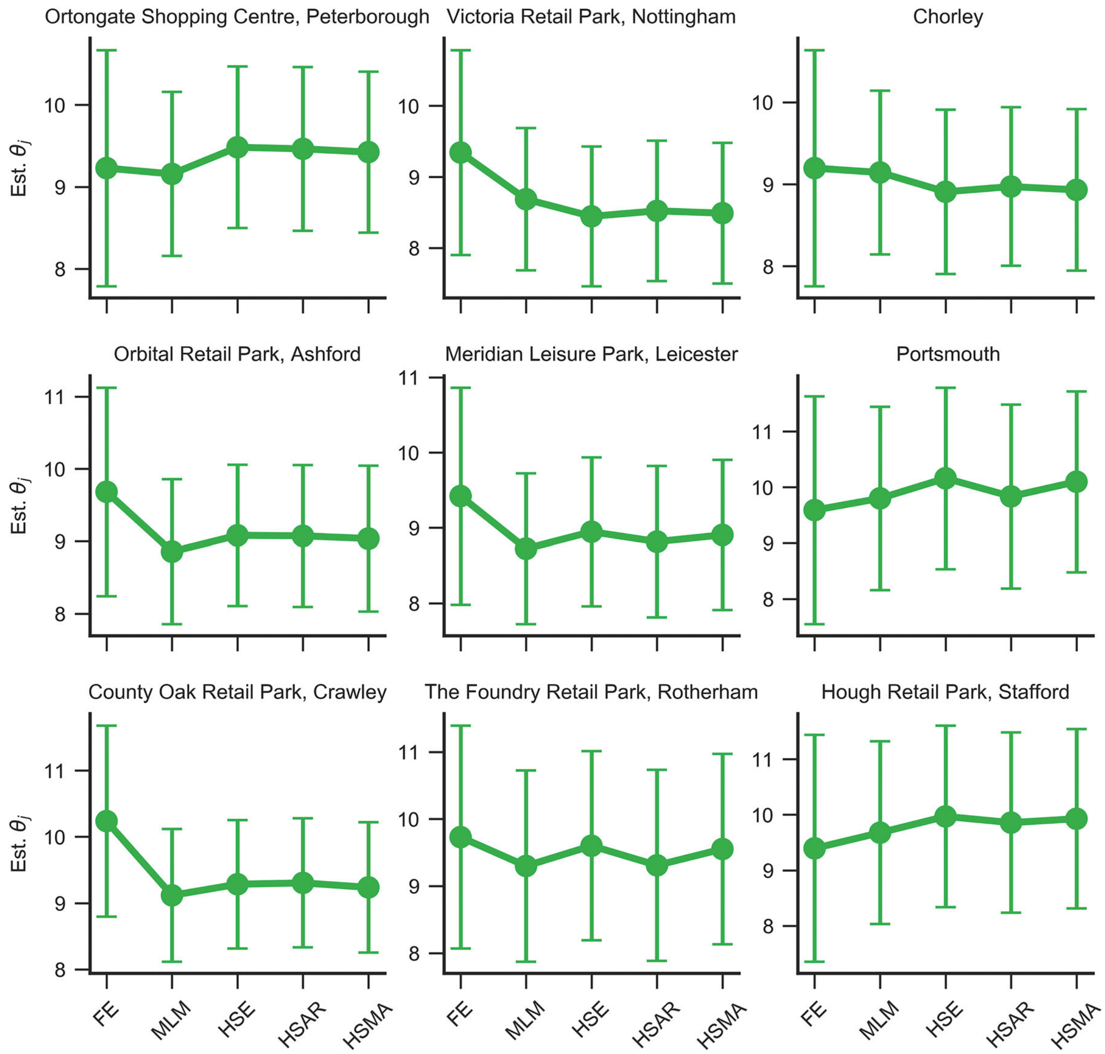


Figure 3. Model estimates for nine retail center boundaries sampled from the retail center willingness to pay rankings. *Note:* FE = fixed effect; MLM = multilevel model; HSE = hierarchical spatial error model; HSAR = hierarchical spatial autoregressive model; HSMA = hierarchical spatial moving average model.

specification(s) over the fixed effect model, with these models able to flexibly accommodate the covariance structure induced by the grouping of stores by retail center boundary. Our search for a preferred specification continues by evaluating potential spatial dependence in the RWTP effect u_j estimated by the MLM. Given that the MLM assumes RWTP values to be independent of each other, we follow G. Dong and Harris (2015) and use a Moran's I to test whether the estimates for RWTP are spatially dependent. A Moran's I statistic for u_j premised on the spatial weights matrix M for the retail center polygons returns a coefficient value of 0.174 ($p > 0.001$). This illustrates positive spatial autocorrelation for the estimated RWTP values, which motivates using the spatial models given that the core model assumption of independence for u_j across retail centers does not hold.

We subsequently turn direct attention to the spatial multilevel models. Given that our hierarchical approach is fully Bayesian, trace plots are required to monitor the convergence of each parameter to the target distribution (see Appendix). In each case the parameters were assessed to have converged. Moreover, there was no serial autocorrelation identified in the stationary Markov chain for each parameter. The first substantive difference we observe is that not accounting for spatial dependence leads the MLM to marginally overestimate the retail center boundary variance σ_u^2 relative to the spatial models; σ_u^2 can be

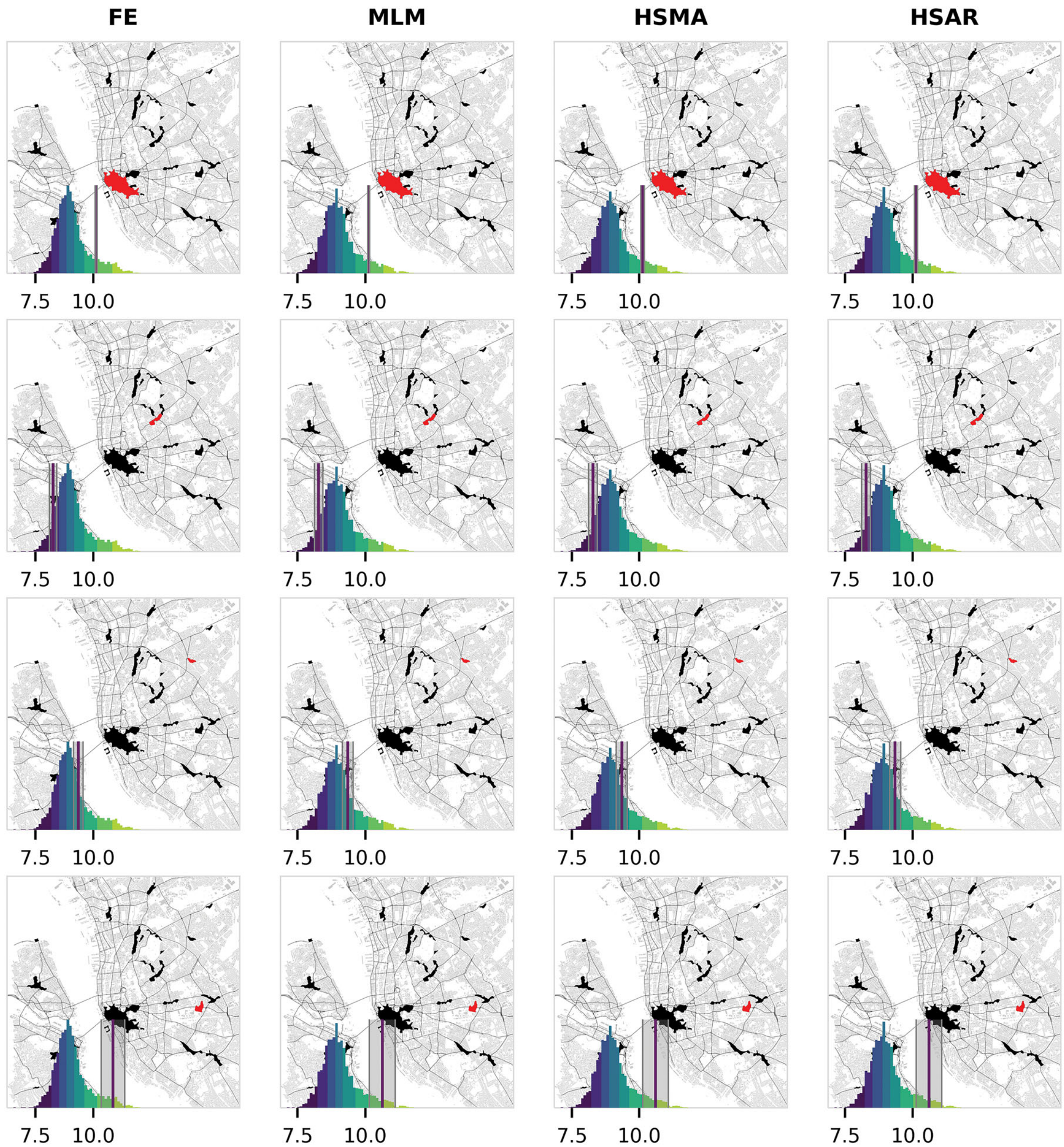


Figure 4. Retail center willingness to pay values in log units for Merseyside, UK. *Note:* Purple vertical bar identifies the value of the retail center highlighted in red, with 95 percent confidence or credibility intervals shaded in gray either side. FE = fixed effect; MLM = multilevel model; HSMA = hierarchical spatial moving average model; HSAR = hierarchical spatial autoregressive model.

understood as the average variation of RWTP values across the retail centers in log units. Here, σ_u^2 falls from 0.503 in the MLM to 0.484, 0.477, and 0.492 in the HSAR, HSE, and HSMA models, respectively. We also recover evidence of a significant spatial autoregressive parameter λ, which is indicative of spatial spillover effects of RWTP values between neighboring retail centers. This is recovered because λ is distinct from zero at the 95 percent credible interval. Interestingly, the density of the covariance structure seems to affect the estimate for λ. The HSMA model, with a sparse covariance structure that is restricted to first- and second-order neighbors, estimates a λ value of 0.189. On the other hand, models with a denser covariance structure such as the HSAR and HSE estimate highly similar values of 0.232 and 0.230. Each of these estimates indicates spatial interaction effects among retail center boundaries.

To aid the visualization of spatial patterning, we illustrate the case of Liverpool in Figure 4 with assistance of legendgrams that show the distribution of RWTP values across all 2,951 retail centers, color coded using $k = 8$ break points classified using Fisher–Jenks optimization (Jenks 1967). Each cell highlights a selected retail center in red, with the corresponding RWTP estimate shown by the vertical bar stemming from the *x*-axis of the legendgram, with 95 percent confidence and credible intervals shaded on either side to highlight uncertainty in the estimate. From left to right, the columns identify the RWTP estimate for the fixed effect, multilevel, HSAR, and HSMA models. From a first reading, the spatial patterning in Figure 4 seems to reveal a fragmented picture of vitality and decline, with less desirable retail centers observed in the immediate hinterland of the prospering regional center (identifiable by the large red polygon in the first row). Overall, from this reading of Figure 4, we are able to discern spatial hierarchies that possibly fragment Merseyside's functional market area, with certain retail centers eliciting a higher willingness to pay than neighboring centers.

Technical Validation

After motivating our preferred methodological approach, we undertake a validation exercise to evaluate whether the estimated RWTP effect θ_j for each retail center in the HSAR model responds to characteristics that are generally identifiable for prospering and thriving areas. Here, we regress θ_j on a selection of variables using ordinary least squares first to assess whether any of the variation in the estimated RWTP values can be attributed to variation in the selected explanatory variables and, second, to quantify the strength of relationship, if any, between the response and explanatory features. Principal attention is paid to the 2011 census Workplace Zone (WZ) population characteristics (Mitchell 2014) that represent individuals working in the retail center. As commuter patterns change, the spatial distribution of the working population changes, which holds when the bulk of economic activity occurs during "traditional" office hours (Mitchell 2014), and WZ statistics are preferable because they describe the daytime working population who commute to their places of work inside the retail center. The WZ variables we use include the percentage of people who report their general health as "good" or better, the percentage of individuals with no qualifications, the percentage of homeowners, the percentage of workers enrolled in higher managerial occupations, and the percentage of individuals in full-time employment. Other variables we consider include the vacancy rate of stores in the retail center calculated from the LDC database[6]; a raw count of stores from the LDC database; the amount of urban green space (m^2; Daras et al. 2018); logged median housing values for the 2015 rolling year (Land Registry 2016); and, finally, binary variables for regions in England and Wales that reflect Nomenclature of Territorial Units for Statistics (NUTS) subdivisions—North West, London, West Midlands, and Wales, for example. In each case, the variables are spatially joined[7] from WZ statistical units to the retail center boundary polygons.

The findings are displayed in Table 2. Generally, they are consistent with expectations, although there are deviations from conventional wisdom. For WZ characteristics, an increase in the number of individuals with "good" health (or better) by 1 percent increases the RWTP value by 3.9 percent. Similarly, an increase in the number of people with no qualifications by 1 percent decreases the value by 2.9 percent. Surprisingly, an increase in the number of workers in higher managerial occupations by 1 percent decreases the RWTP of the retail center by 4.9 percent. At first glance this result appears counterintuitive, but managerial workers are more likely to work in financial districts characterized by mostly office space, which are not necessarily perceived as desirable in the same way that consumer amenities such as leisure plazas and urban green spaces are.

Next, we consider retail center boundary characteristics. For every additional 100 stores in the retail center, the RWTP value increases by 4.3 percent, which implies that patrons value a large number of available retail destinations. Similarly, as the vacancy rate increases by 1 percent, the RWTP of the area decreases by 2.2 percent. Again, this is consistent with expectations that a large number of vacant units deteriorates the vibrancy of the streetscape by revealing signs of decay. On the other hand, the availability of urban green space was not a significant determinant. For the regional indicators, relative to the East Midlands reference category, we recover some examples of regional inequality. Whereas retail centers in the East of England are estimated as having the highest RWTP value (28.9 percent), there is a clear disparity in the estimated values for North West England (9.5 percent), South West England (4.1 percent), and to a lesser extent, Yorkshire and

Table 2. Regression results for validation exercise

	Dependent variable HSAR θ_J (1)
(Intercept)	2.684***
	(0.485)
Workplace zone characteristics	
Good health (%)	0.039***
	(0.002)
No qualifications (%)	−0.029***
	(0.005)
Tenure owned (%)	0.0003
	(0.001)
Higher managerial occupations (%)	−0.049***
	(0.003)
Full-time work (%)	0.020***
	(0.001)
Retail center characteristics	
Vacancy rate	−0.017***
	(0.002)
Store count	0.024***
	(0.004)
Urban green space	−0.002
	(0.004)
ln median house price 2015	0.243***
	(0.039)
Regional indicators	
East of England	0.289***
	(0.056)
London	0.203***
	(0.061)
North East England	0.095
	(0.067)
North West England	0.023
	(0.053)
South East England	0.203***
	(0.052)
South West England	0.041
	(0.054)
Wales	−0.066
	(0.068)
West Midlands	0.229***
	(0.057)
Yorkshire and the Humber	0.135***
	(0.054)
Observations	2,951
RMSE	0.555
Adjusted R^2	0.352

Notes: Regions reference category is East Midlands. HSAR = hierarchical spatial autoregressive model; RMSE = root mean square error.
$^{*}p < 0.1$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

The Humber (13.5 percent) when compared to South East England (20.3 percent) and London (20.3 percent). These inequalities are broadly consistent with regional variations in wealth across England and Wales (Rowlingson and McKay 2011). In all, the validation exercise demonstrates a relationship between RWTP values and socioeconomic characteristics that is consistent with conventional wisdom. Although not conclusive, the coefficients of our estimates suggest that a decline in RWTP is related to urban environments with poorer social and community well-being. This begins to address a key gap in the evidence linking retail center outcomes to characteristics of the urban environment that is identified by the Department of Business, Innovation and Skills for England and Wales (BIS 2011).

Conclusion

The depth and breadth of leisure and retail opportunity are increasingly linked to the desirability of places to live (Glaeser, Kolko, and Saiz 2001). Because the quality of urban environments cannot be qualified by a natural unit of analysis, the willingness to pay to receive an amenity-rich environment has often been explored through the lens of the residential housing market. The groundings of this article were motivated by similar hedonic analyses, except that we used business rates for commercial properties alongside a nontrivial methodological framework to estimate RWTP, for which we provide a detailed exposition for reproducing the analysis. Similar to approaches that analyze housing prices, by controlling for property-level characteristics such as the total floor area, car parking spaces, and store type, the remaining variation in the business rate was attributed to the RWTP. This was possible because business rates approximate local market conditions, because ratable values are set by estimating a basic cost per square meter that is adjusted to reflect similar properties in the same area (VOA 2014). Despite our empirical motivations, particular attention to how the RWTP estimates interface with the unique geographic behavioral characteristics of the UK retail landscape was required. Due to restructuring of the traditional brick-and-mortar retailer landscape through growth in electronic retailing, our study required particular attention to the nuances of UK retail spaces. It is often argued that growth in online retailing is forecast by its deleterious effects that cause physical shopping opportunity to be substituted online (Doherty and Ellis-Chadwick 2010). Despite these concerns, online retail has recently been linked to complementarity and modification processes. These processes blend traditional retail with e-commerce through integration of technologies such as click and collect points that operate as points of delivery for Internet sales (Singleton et al. 2016). Thus, through the market system of using business rates, the RWTP estimates relate to how much the behavior of consumers values a given retail area. Among the context of behavioral patterns, this allowed us to unpack hierarchies of retail spaces. These spaces are an underlying driver to the sustainability of built environments and so, by implication, reveal geographical patterns in urban growth and development.

Multilevel models have a rich history in the educational sciences literature for building league tables of school performance (Goldstein 2003). We used similar motivations to build a ranking of retail centers, except that unlike previous studies, we allowed for possible spatial autocorrelation that operates on the basis of geographical proximity. This is because the RWTP effect per retail center is likely to covary based on spatial proximity. With these motivations, and by revamping the traditional focus of multilevel modeling techniques, we were able to derive retail center estimates of RWTP. A particular focus on retail centers, our geography of choice, was because they have been argued as a moderating influence on urban hierarchies (Dennis, Marsland, and Cockett 2002). Yet, there is a limited availability of national data for measuring the economic and social value of retail centers, with a presumptive attitude in UK policy circles that the impacts of policy instruments such as the Town Centers First approach are "instinctively positive" (BIS 2011). In producing ranked estimates, we remedied these uncertainties by building quantifiable evidence to directly observe disparities in RWTP across networks of retail centers. More concretely, the derived scores allow an understanding of a particular retail center's position within a network of centers; this can be used as a proxy of economic health and an indicator of the pull that particular retail center catchments have on consumers in the area. From this, retail practitioners might be able to use the derived scores as proxies for footfall generation, which would allow them to deduce consumer appeal of particular centers. Knowledge of such characteristics might be used in decision-making processes, such as determining investment and divestment outcomes or the rationalization of store portfolios, for example. At a general level, our findings also provide a platform for researchers to build on. The applied methodology provides a blueprint for constructing hierarchies of retail centers that is replicable and generalizable to similar contexts, conditional on data availability. In addition, to our knowledge, the study is the first of its kind to build indicators that describe hierarchies of retail centers across a national extent, with previous studies typically limited to smaller case study areas. Finally, a core and intentional contribution of the article is the potential for exploration of hypotheses in retail geography that were previously unavailable due to the absence of statistical data on retail centers.

To conclude this article, we illustrate elaborations to consider for future research. One refinement involves the addition of further attributes at the store or retail center level to be specified into the modeling approach. This might involve undertaking visual, in-person surveys for small case study areas to collect image attributes identified in Gomes and Paula (2017) such as parking security, atmosphere perception, or mix and quality of stores within the retail center boundary, for example. Due to the practicality concerns of obtaining these highly granular measures in this study, this direction would reduce the number of retail centers for which the approach can return RWTP estimates. The benefit, however, is that it would allow an estimation of the willingness to pay for highly granular measures that describe image-based attributes of attractive shopping environments. As a final remark, the advantage of the applied methodology is that it can be redeployed in the future to generate timely updates. This is possible because the VOA continues to reassess the ratable values of nondomestic properties according to a five-year revaluation cycle (VOA 2014). Conditional on the VOA continuing to release their ratings list as an open data product, the area estimates of RWTP are updatable over time. Future research might develop retail center rankings into a longitudinal data product that allows an exploration into the temporal characteristics of RWTP and how successive five-year windows alter the rank-ordered positions of retail centers.

Funding

This project was funded by ESRC studentship funding.

ORCID

Daniel Arribas-Bel http://orcid.org/0000-0002-6274-1619

Notes

1. Given this article's intersection between retail geography and urban economics, particular attention to the conceptualization of attractiveness is required. Whereas urban economists perceive attractiveness through an estimation of willingness to pay, retail geographers might observe the attractiveness of shopping environments through a lens of image-based characteristics such as cleanliness of the shopping environment, plurality and variety of shops, or existence of fun and entertainment programs (El-Adly 2007; Chebat, Sirgy, and Grzeskowiak 2010; Gomes and Paula 2017). Thus, to avoid confusion, in this article we adopt the direction of the former and describe our measure of interest by willingness to pay.
2. Spatial connectivity at the retail center level is specified as

$$M_{ij} = \begin{cases} 1, \exp\left(-\left(d_{ij}^2\right)/d^2\right), if d_{ij} \leq 0 \\ 0, \text{otherwise} \end{cases},$$

   where d_{ij} is the Euclidean distance between retail center and d is the fixed-distance bandwidth. A semivariogram was used as an exploratory tool for determining the distance at which the spatial dependence between business rates between retail centers became negligible (see Appendix).
3. The following conjugate priors are chosen:

$$P(\beta) \propto \mathcal{N}(0, 100)$$
$$P\left(\sigma_e^2\right) \propto \text{InverseGamma}(0.01, 0.01)$$
$$P\left(\sigma_u^2\right) \propto \text{InverseGamma}(0.01, 0.01)$$
$$P(\lambda) \propto \text{Uniform}(-1, 1).$$

4. LDC premise types were recoded in accordance with VOA Special Categories outlined in Rhodes and Brien (2017).
5. Potential problems of multicollinearity were assessed using variance inflation factor (VIF) scores for each predictor variable in the spatial fixed effect model. VIF scores revealed no evidence of such problems, with scores of about 3.0 leading us to continue with our inferential exercise.
6. Vacancy rates are defined as the proportion of all available retail units that are vacant or unoccupied.
7. Because there is only partial overlap between the retail centers and WZ polygons, the resulting WZ statistics are aggregated by the mean value for the intersecting WZ geometries when joined to each retail center polygon.

References

Adam, S., and H. Miller. 2014. Business rates. Accessed June 21, 2019. http://www.ifs.org.uk/budgets/gb2014/gb2014_ch11.pdf.

Anselin, L. 1988. *Spatial econometrics: Methods and models*. Dordrecht, The Netherlands: Kluwer Academic.

Anselin, L. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26 (2):153–66. doi: 10.1177/0160017602250972.

Anselin, L., and D. Arribas-Bel. 2013. Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science* 92 (1):3–17. doi: 10.1111/j.1435-5957.2012.00480.x.

Astbury, G., and M. Thurstain-Goodwin. 2014. Measuring the impact of out-of-town retail development on town centre retail property in England and Wales. *Applied Spatial Analysis and Policy* 7 (4):301–16. doi: 10.1007/s12061-014-9111-5.

Birkin, M., G. Clarke, and M. Clarke. 2002. *Retail geography and intelligent network planning*. Chichester, UK: Wiley.

Browne, W., S. Subramanian, K. Jones, and H. Goldstein. 2005. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168 (3):599–613. doi: 10.1111/j.1467-985X.2004.00365.x.

Brueckner, J., J. Thisse, and Y. Zenou. 1999. Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review* 43 (1):91–107. doi: 10.1016/S0014-2921(98)00019-1.

CACI. 2018. Retail footprint—CACI's leading retail centre catchment model. Accessed June 21, 2019. https://www.caci.co.uk/products/product/retail-footprint.

Chalmers, K., M. Gessner, L. Venturoni, and S. Weiler. 2012. Stemming retail leakage with a sense of community: Leveraging the links between communal ties and shopping decisions. *The Social Science Journal* 49 (1):108–13. doi: 10.1016/j.soscij.2011.06.010.

Chebat, J., J. Sirgy, and S. Grzeskowiak. 2010. How can shopping mall management best capture mall image? *Journal of Business Research* 63 (7):735–40. doi: 10.1016/j.jbusres.2009.05.009.

Cho, S., and J. Workman. 2011. Gender, fashion innovativeness and opinion leadership, and need for touch: Effects on multi-channel choice and touch/non-touch preference in clothing shopping. *Journal of Fashion Marketing and Management: An International Journal* 15 (3):363–82. doi: 10.1108/13612021111151941.

Clark, N. 2003. *The city as an entertainment machine*. Bingley, UK: Emerald Group.

Daras, K., A. Davies, M. Green, and A. Singleton. 2018. Developing indicators for measuring health-related features of neighbourhoods. In *Consumer data research*, ed. P. A. Longley, A. Singleton, and J. Cheshire, 167–77. London: UCL Press.

Dennis, C., D. Marsland, and T. Cockett. 2002. Central place practice: Shopping centre attractiveness measures, hinterland boundaries and the UK retail hierarchy. *Journal of Retailing and Consumer Services* 9 (4):185–99. doi: 10.1016/S0969-6989(01)00021-2.

Department of Business, Innovation and Skills. 2011. Understanding high street performance. Accessed June 21, 2019. http://urbanpollinators.co.uk/wp-content/plugins/downloads-manager/upload/understanding high-street-performance.pdf.

Department of the Environment Urban and Economic Development Group. 1994. *Vital and viable town centres: Meeting the challenge*. London: Department of the Environment Urban; Economic Development Group. Accessed June 21 2019. http://www.nrpf.org.uk/PDF/Vital%20and%20Viable%20Town%20Centres%20-%20%20DoE%20(1994).pdf.

Doherty, N., and F. Ellis-Chadwick. 2010. Internet retailing: The past, the present and the future. *International Journal of Retail & Distribution Management* 38 (11–12):943–65. doi: 10.1108/09590551011086000.

Dolega, L., M. Pavlis, and A. Singleton. 2016. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services* 28:78–90. doi: 10.1016/j.jretconser.2015.08.013.

Dong, G., and R. Harris. 2015. Spatial autoregressive models for geographically hierarchical data structures. *Geographical Analysis* 47 (2):173–91. doi: 10.1111/gean.12049.

Dong, G., and W. Wu. 2016. Schools, land markets and spatial effects. *Land Use Policy* 59:366–74. doi: 10.1016/j.landusepol.2016.09.015.

El-Adly. M. 2007. Shopping malls attractiveness: A segmentation approach. *International Journal of Retail & Distribution Management* 35 (11):936–50. doi: 10.1108/09590550710828245.

Ethical Consumer Research Association. 2013. Ethical consumer market report. Accessed June 21, 2019. https://www.ethicalconsumer.org/sites/default/files/inline-files/ethical_consumer_markets_report_2013.pdf.

Finn, A., and J. Louviere. 1996. Shopping center image, consideration, and choice: Anchor store contribution. *Journal of Business Research* 35 (3):241–51. doi: 10.1016/0148-2963(95)00129-8.

Florida, R. 2000. Competing in the age of talent: Quality of place and the new economy. Accessed June 21 2019. https://www.creativeclass.com/rfcgdb/articles/2000-Competing_In_The_Age_Of_Talent.pdf.

Florida, R. 2008. *Who's your city? How the creative economy is making where you live the most important decision of your life*. New York: Random House.

Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. *Analytical methods for social research*. New York: Cambridge University Press.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (6):721–41. doi: 10.1109/TPAMI.1984.4767596.

Glaeser, E., J. Kolko, and A. Saiz. 2001. Consumer city. *Journal of Economic Geography* 1 (1):27–50. doi: 10.1093/jeg/1.1.27.

Goldstein, H. 2003. Multilevel modelling of educational data. In *Methodology and epistemology of multilevel analysis: Approaches from different social sciences*, ed. D Courgeau, 25–42. Dordrecht, The Netherlands: Springer.

Goldstein, H. 2011. *Multilevel statistical models*. West Sussex, UK: Wiley.

Gomes, R., and F. Paula. 2017. Shopping mall image: Systematic review of 40 years of research. *The International Review of Retail, Distribution and Consumer Research* 27 (1):1–27. doi: 10.1080/09593969.2016.1210018.

Goss, J. 1993. The "magic of the mall": An analysis of form, function, and meaning in the contemporary retail built environment. *Annals of the Association of American Geographers* 83 (1):18–47. doi: 10.1111/j.1467-8306.1993.tb01921.x.

Helm, S., S. Kim, and V. Silvia. 2018. Navigating the retail apocalypse: A framework of consumer evaluations of the new retail landscape. *Journal of Retailing and Consumer Services*. Advance online publication. doi: 10.1016/j.jretconser.2018.09.015.

Helsper, E., and R. Eynon. 2010. Digital natives: Where is the evidence? *British Educational Research Journal* 36 (3):503–20. doi: 10.1080/01411920902989227.

Huff, D. 1963. A probabilistic analysis of shopping center trade areas. *Land Economics* 39 (1):81–90. doi: 10.2307/3144521.

Hui, E., and C. Liang. 2016. Spatial spillover effect of urban landscape views on property price. *Applied Geography* 72:26–35. doi: 10.1016/j.apgeog.2016.05.006.

Institute of Grocery Distribution. 2014. Shopper trends. Accessed June 21, 2019. http://www.igd.com/our-expertise/Shopper-Insight/shopper-outlook/4893/Shopper-Trends.

Javelin Group. 2017. Venuescore 2017—UK shopping venue rankings. Accessed June 21, 2019. https://1tdmn8g702e2kt13m2s6be22-wpengine.netdna-ssl.com/wp-content/uploads/2017/01/Javelin_Group_Executive_Summary_VENUESCORE_2017-1.pdf.

Jenks, G. 1967. The data model concept in statistical mapping. *International Yearbook of Cartography* 7:186–90.

Jin, B., and B. Sternquist. 2004. Shopping is truly a joy. *The Service Industries Journal* 24 (6):1–18. doi: 10.1080/0264206042000299158.

Jones, K. 1991. Specifying and estimating multilevel models for geographical research. *Transactions of the Institute of British Geographers* 16 (2):148–59. doi: 10.2307/622610.

Land Registry. 2016. Price paid open data. Accessed June 21, 2019. http://landregistry.data.gov.uk/.

Lesage, J. 1997. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20 (1–2):113–29. doi: 10.1177/016001769702000107.

Lesage, J. 2014. What regional scientists need to know about spatial econometrics. *The Review of Regional Studies* 44 (1):13–32.

Longley, P., and A. Singleton. 2009. Linking social deprivation and digital exclusion in England. *Urban Studies* 46 (7):1275–98. doi: 10.1177/0042098009104566.

Mitchell, B. 2014. Workplace zones: A new geography for workplace statistics. Office for National Statistics. Accessed June 21, 2019. https://data.gov.uk/dataset/workplace-zones-a-new-geography-for-workplace-statistics3.

Office for National Statistics. 2018. Internet access—Households and individuals, Great Britain

2018. Accessed June 21, 2019. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018.
Oner, O. 2017. Retail city: The relationship between place attractiveness and accessibility to shops. *Spatial Economic Analysis* 12 (1):72–91. doi: 10.1080/17421772.2017.1265663.
Owen, G., R. Harris, and K. Jones. 2016. Under examination: Multilevel models, geography and health research. *Progress in Human Geography* 40 (3):394–412. doi: 10.1177/0309132515580814.
Pavlis, M., L. Dolega, and A. Singleton. 2017. A modified DBSCAN clustering method to estimate retail center extent. *Geographical Analysis* 50:141–61. doi: 10.1111/gean.12138.
Poushneh, A., and A. Vasquez-Parraga. 2017. Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. *Journal of Retailing and Consumer Services* 34:229–34. doi: 10.1016/j.jretconser.2016.10.005.
Rappaport, J. 2009. The increasing importance of quality of life†. *Journal of Economic Geography* 9 (6):779–804. doi: 10.1093/jeg/lbp009.
Reilly, W. 1931. *The law of retail gravitation*. New York: Knickerbocker.
Rhodes, C., and P. Brien. 2017. The retail industry: Statistics and policy. Accessed June 21, 2019. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06186.
Rivera-Batiz, F. 1988. Increasing returns, monopolistic competition, and agglomeration economies in consumption and production. *Regional Science and Urban Economics* 18 (1):125–53. doi: 10.1016/0166-0462(88)90008-7.
Rosen, S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82 (1):34–55. doi: 10.1086/260169.
Rowlingson, K., and S. McKay. 2011. *Wealth and the wealthy: Exploring and tackling inequalities between rich and poor*. Bristol, UK: Bristol University Press.
Sack, R. 1988. The consumer's world: Place as context. *Annals of the Association of American Geographers* 78 (4):642–64. doi: 10.1111/j.1467-8306.1988.tb00236.x.
Singleton, A., L. Dolega, D. Riddlesden, and P. Longley. 2016. Measuring the spatial vulnerability of retail centres to online consumption through a framework of e-resilience. *Geoforum* 69:5–18. doi: 10.1016/j.geoforum.2015.11.013.
Teller, C., and J. Elms. 2010. Managing the attractiveness of evolved and created retail agglomerations formats. *Marketing Intelligence & Planning* 28 (1):25–45. doi: 10.1108/02634501011014598.
Teller, C., and J. Elms. 2012. Urban place marketing and retail agglomeration customers. *Journal of Marketing Management* 28 (5–6):546–67. doi: 10.1080/0267257X.2010.517710.
Teller, C., and T. Reutterer. 2008. The evolving concept of retail attractiveness: What makes retail agglomerations attractive when customers shop at them? *Journal of Retailing and Consumer Services* 15 (3):127–43. doi: 10.1016/j.jretconser.2007.03.003.
Thurstain-Goodwin, M., and D. Unwin. 2000. Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. *Transactions in GIS* 4 (4):305–17. doi: 10.1111/1467-9671.00058.
Valuation Office Agency. 2014. VOA rating list downloads. Accessed June 21, 2019. https://www.gov.uk/guidance/how-non-domestic-property-including-plant-and-machinery-is-valued.
Valuation Office Agency. 2018. How non-domestic (business) properties are valued. Accessed June 21, 2019. https://voaratinglists.blob.core.windows.net/html/rlidata.htm.
Warnaby, G., D. Bennison, B. Davies, and H. Hughes. 2002. Marketing UK towns and cities as shopping destinations. *Journal of Marketing Management* 18 (9–10):877–904. doi: 10.1362/0267257012930402.
Warren, M. 2007. The digital vicious cycle: Links between social disadvantage and digital exclusion in rural areas. *Telecommunications Policy* 31 (6–7):374–88. doi: 10.1016/j.telpol.2007.04.001.
Williams, D. 2009. The evolution of e-tailing. *The International Review of Retail, Distribution and Consumer Research* 19 (3):219–49. doi: 10.1080/09593960903233657.
Wolf, L., L. Anselin, D. Arribas-Bel, and L. Mobley. 2018. Local shrinkage in spatially-correlated variance components models. Unpublished preprint, University of Bristol, UK.
Wrigley, N., and L. Dolega. 2011. Resilience, fragility, and adaptation: New evidence on the performance of UK high streets during global economic crisis and its policy implications. *Environment and Planning A: Economy and Space* 43 (10):2337–63. doi: 10.1068/a44270.
Wrigley, N., and D. Lambiri. 2014. High street performance and evolution—A brief guide to the evidence. Accessed June 21, 2019. https://thegreatbritishhigh-street.co.uk/pdf/GBHS-High StreetReport.pdf.
Wrigley, N., D. Lambiri, G. Astbury, L. Dolega, C. Hart, C. Reeves, M. Thurstain-Goodwin, and S. Wood. 2015. British high streets: From crisis to recovery? A comprehensive review of the evidence. Accessed June 21, 2019. https://eprints.soton.ac.uk/375492/.
Zentner, A., M. Smith, and C. Kaya. 2013. How video rental patterns change as consumers move online. *Management Science* 59 (11):2622–34. doi: 10.1287/mnsc.2013.1731.

SAM COMBER is a PhD student in Geographic Data Science at the Department of Geography and Planning, University of Liverpool, Liverpool L69 7ZT, UK. E-mail: s.comber@liverpool.ac.uk. His core research interests include spatial statistics and machine learning with a focus on computational methods for understanding urban areas.

DANIEL ARRIBAS-BEL is a Senior Lecturer in Geographic Data Science at the Department of Geography and Planning, University of Liverpool, Liverpool L69 7ZT, UK. E-mail: D.Arribas-Bel@liverpool.ac.uk. His research interests combine urban studies, computational methods, and new forms of data.

ALEX SINGLETON is a Professor of Geographic Information Science at the University of Liverpool, Deputy Director of the ESRC Consumer Data Research Center (CDRC) and Director of the ESRC Data Analytics & Society CDT, University of Liverpool, Liverpool L69 7ZT, UK. E-mail: alex.singleton@liverpool.ac.uk. His research is concerned with how the complexities of individual behaviors, attitudes, and contexts manifest spatially and can be represented and understood through a framework of geographic data science. He is particularly interested in applications of area classification to understand multidimensional sociospatial structure and has developed a broad critique of the ways in which such geodemographic methods can be refined through modern scientific approaches of machine learning, GIScience, and quantitative human geography.

GUANPENG DONG is a Lecturer in Geographic Data Science at the Department of Geography and Planning, University of Liverpool, Liverpool L69 7ZT, UK. E-mail: guanpeng.dong@liverpool.ac.uk. His core research interests include the development of spatial/spatiotemporal statistical and multilevel modeling approaches and the applications in urban studies.

LES DOLEGA is a Lecturer in the Department of Geography and Planning, University of Liverpool, Liverpool L69 7ZT, UK. E-mail: L.Dolega@liverpool.ac.uk. His interest and expertise are in economic/retail geography and GIS, in particular the performance of town centers and their evolutionary trajectories. Currently, he is working on automated delineation of town center boundaries and a dynamic and multidimensional classification of retail and consumption spaces in the United Kingdom.

Appendix

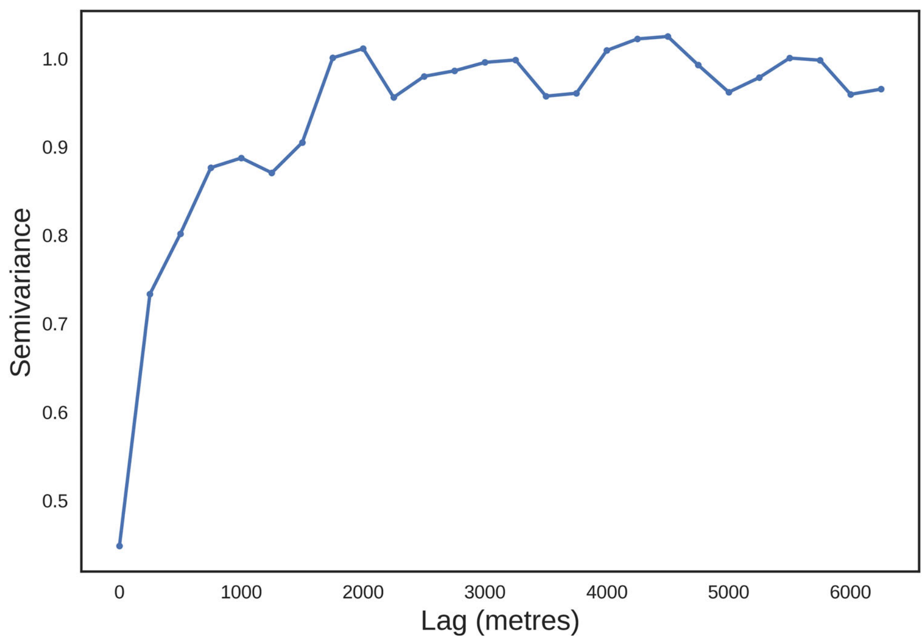


Figure A.1. Semivariogram demonstrating the tendency for retail centers close together in space to exhibit higher correlations for business rates than those further apart.

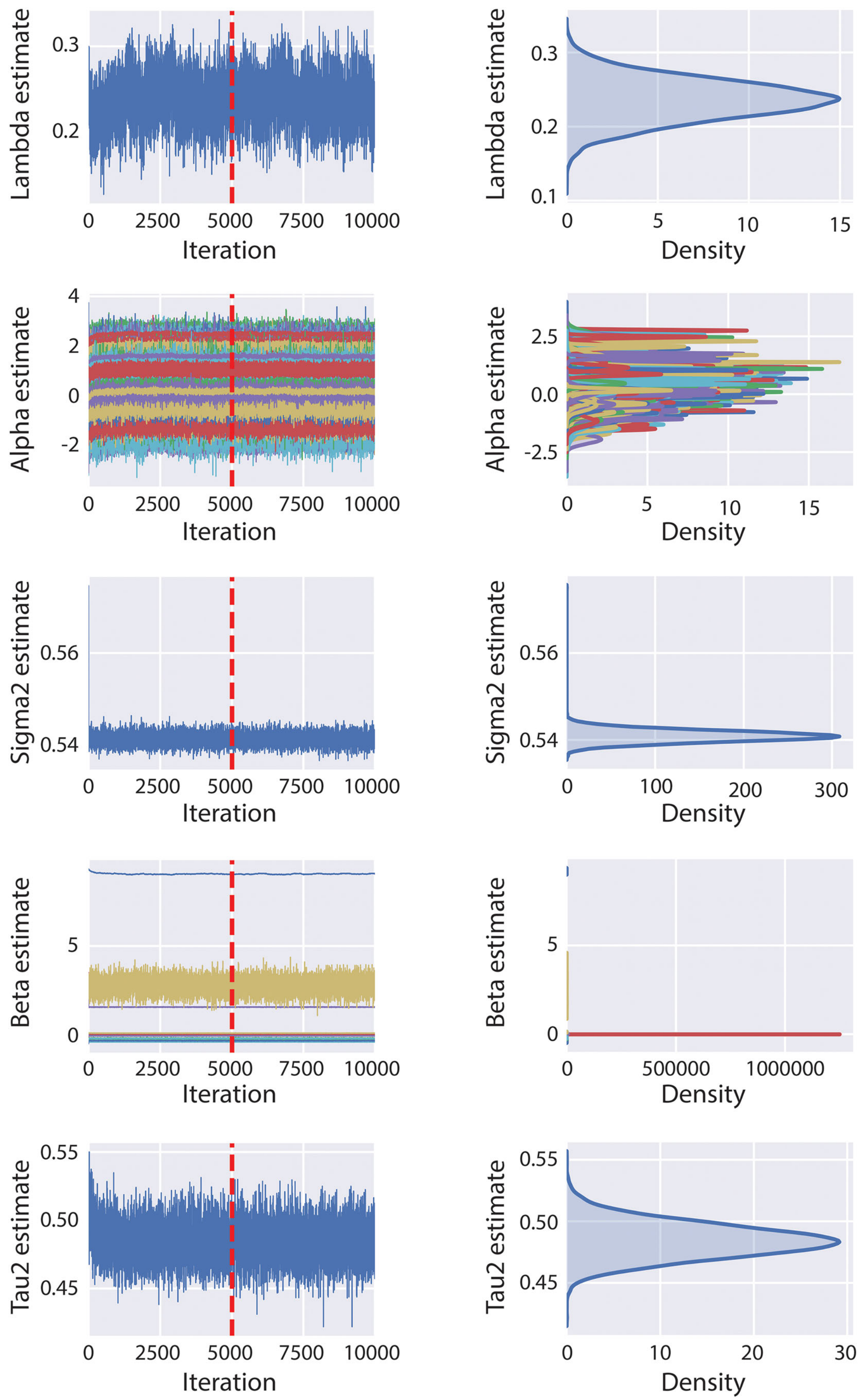


Figure A.2. Trace plots (left) displaying simulated draws from Markov chain Monte Carlo chain for parameters and distribution of samples (right) for hierarchical spatial autoregressive model. Vertical red line highlights the point along the chain where previous samples are discarded.

Table A.1. Variable description for property-level characteristics

Variable	Description	Source	M	SD	Unit
Dependent variable					
Business rate	Ratable value taxed on the business property	VOA	100,639.9	976,771.5	Pounds
Structural characteristics					
Floor area	Total floor area of property (in thousands)	VOA	0.658	6.155	m^2
No. rooms	Number of surveyable rooms	VOA	6.577	4.010	Count
Parking	Number of car parking spaces	VOA	0.099	2.053	Count
Store typology					
Banks and other A2 uses	1 for A2 uses, 0 otherwise	LDC	0.07	0.25	Binary
Factory shops	1 for factory shops, 0 otherwise	LDC	0.004	0.07	Binary
Food stores	1 for food store (<750 m^2), 0 otherwise	LDC	0.04	0.20	Binary
Hairdressing/beauty salon	1 for salon, 0 otherwise	LDC	0.13	0.33	Binary
Hypermarket/superstore	1 for superstore ($>2,500$ m^2), 0 otherwise	LDC	0.03	0.16	Binary
Large food stores	1 for large food store (>750 m^2), 0 otherwise	LDC	0.04	0.20	Binary
Large retail shops	1 for large shops ($>1,850$ m^2), 0 otherwise	LDC	0.04	0.20	Binary
Nonretail	1 for nonretail, 0 otherwise	LDC	0.05	0.21	Binary
Other	1 for other premises, 0 otherwise	LDC	0.01	0.12	Binary
Pharmacies	1 for pharmacy, 0 otherwise	LDC	0.004	0.06	Binary
Post offices	1 for post office, 0 otherwise	LDC	0.01	0.10	Binary
Restaurants and bars	1 for restaurant or bar, 0 otherwise	LDC	0.07	0.25	Binary
Retail shops	1 for high street retail store, 0 otherwise	LDC	0.40	0.49	Binary
Takeaway food outlet	1 for takeaway outlet, 0 otherwise	LDC	0.11	0.32	Binary
Showrooms	1 for showroom, 0 otherwise	LDC	0.03	0.18	Binary

Notes: VOA = Valuation Office Agency; LDC = Local Data Company.

C — Landscape and Urban Planning paper

Landscape and Urban Planning 202 (2020) 103887

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Using convolutional autoencoders to extract visual features of leisure and retail environments

Sam Comber*, Daniel Arribas-Bel, Alex Singleton, Les Dolega

Department of Geography & Planning, University of Liverpool, Roxby Building, Liverpool L69 7ZT, United Kingdom

A B S T R A C T

Visual characteristics of leisure and retail environments provide sensory cues that can influence how consumers experience and behave within these spaces. In this paper, we provide a computational method that summarises the "visual features" of shopping districts by analysing a national database of geocoded store frontage images. While the traditional focus of social scientific research explores how drivers such as proximity to shopping environments factor into location choice decisions, the visual characteristics that describe the enclosing urban area are often neglected. This is despite the assumption consumers translate visual appearance of a retail area into a judgement of its functional utility which mediates consumer behaviour, patronage intention and the image a retail location projects to passers-by. Such judgements allow consumers to draw fine distinctions when evaluating between competing destinations. Our approach introduces a deep learning model known as Convolutional Autoencoders to extract visual features from storefront images of leisure and retail amenities. These features are partitioned into five clusters before several measures describing the environment around the leisure and retail properties are introduced to differentiate between the clusters and assess which variables are distinctive for particular groupings. Our empirical strategy unpacks different groupings from the clusters, which implies the existence of relationships between visual features of shopping areas and functional characteristics of the surrounding urban environment. Ultimately, using the example of retail landscapes, the core contribution of this paper demonstrates the utility of unsupervised deep learning methods to research questions in urban planning.

1. Introduction

Visual characteristics of urban spaces drive how individuals evaluate and experience their surroundings for the purpose of location choice behaviour and patronage decisions (Hauser & Koppelman, 1979). In the *The Image of the City*, Kevin Lynch argues the built environment can be drawn as "mental maps" that describe how the city is read visually by cues such as shapes, sizes and colours (Lynch, 1960). Not only this, Silver and Clark (2016) argue the actions, tastes, and traits of individuals create and support particular meanings attached to places. The measurement of a scene assesses the character of a particular place and highlights distinctive visual aspects of the built environment. As visual (but subjective) measures that describe scenes such as liveliness are hard to quantify with traditionally-available data, urban planners typically resort to building indicators that are based on more directly observable characteristics such as population density (Glaeser and Gottlieb, 2009) or street layout (Jung et al., 2017). Often representations that characterise the scenes of streets are inferred using visual audits conducted by researchers who collate data to explore similarities and differences of physical attributes visible from street-level – the quality of building facade, the presence of street art, or the condition of sidewalks, for example (Bader et al., 2017). Once aggregated, researchers can unpack relationships exploring the link between particular visual attributes of built environments and characteristics of the surrounding area. For retail environments, the visual image that shopping areas project to consumers is a function of a broad range of influences which affect patronage behaviour and consumer experiences (Bell, 1999). Retail area image is a multidimensional concept and to understand it requires unpacking the multitude of functional and visual characteristics that consumers associate with shopping areas (Baker et al., 1994). These characteristics are stimuli that influence consumer perception and, by extension, patronage intention for particular retail environments. Typically measures of retail area image are derived using survey approaches that rate characteristics such as the quality of building materials, the attractiveness of shop signage and overall environmental cleanliness (Bellizzi et al., 1983; El-Adly, 2007). As conducting in-person studies in shopping areas to record this data requires a high level of human judgement, they are cost-intensive and limited in the throughput required to construct visual descriptors of retail environments for large study areas

To circumvent the scalability issues of manually auditing a national sample of retail locations, we apply Convolutional Autoencoders (CAEs) to automatically extract visual features from images showing the frontage of leisure and retail properties across England and Wales. Particular interest on street-level imagery for leisure and retail amenities stems from their influence to the vibrancy of places and, hence, in

* Corresponding author.

E-mail addresses: s.comber@liverpool.ac.uk (S. Comber), D.Arribas-Bel@liverpool.ac.uk (D. Arribas-Bel), Alex.Singleton@liverpool.ac.uk (A. Singleton), L.Dolega@liverpool.ac.uk (L. Dolega).

https://doi.org/10.1016/j.landurbplan.2020.103887

Received 8 November 2019; Received in revised form 15 June 2020; Accepted 21 June 2020

Available online 11 July 2020

the characteristics of the urban hierarchy (Dennis et al., 2002). While previous studies have shown that proximity to leisure and retail amenities factor into location choice decisions and patronage of retail environments (Glaeser and Gottlieb, 2009), the visual characteristics that describe the urban environment around the point of interest are neglected. Such approaches assume a "vacuum" around single amenities, which ignores the environmental context that surrounds these premises. As an example, the visual characteristics of a street with a restaurant accessible by several modes of transportation is likely to differ by the amount of liveliness when compared with another restaurant serviced in a location with no transport links. Capturing visual features of leisure and retail amenities allows an exploration into whether aspects of *what we see* are related to particular characteristics of the built environment that describe the amenities location. By clustering visual features extracted from the CAE, the principal contribution of this paper uses deep learning to assess whether visual-only features of retail landscapes correlate with observed characteristics of built environments, and whether there are distinctive characteristics for particular groupings.

The remainder of the paper is organised as follows. Section 2 motivates the underlying conceptual framework of the paper. Section 3 introduces the sources of data we utilise through the study, before describing the modelling approach we implement to arrive at our empirical objective. Section 4 presents the main findings. Finally, Section 5 concludes the paper.

2. Background and motivation

2.1. Visual characteristics of built environments

In the *Critique of Judgement* Immanuel Kant first observed aesthetic perception as a self-organising process that drives how individuals react to different environments (Kant, 1790). Not only do humans perceive their environment as neutral facts and data, but we react to distinctive aesthetic cues encoded in our surroundings that change how these spaces are experienced as we walk through them (Silver and Clark, 2016). Our judgement of the elements in our surroundings are rendered as a totality, independent of the constituent parts. When we stroll through a "hip neighbourhood", the avant-garde feel, boutique stores, and DIY atmosphere are not perceived as independent objects. This is because they collectively recall a particular way of behaving that is adopted from the tastes and preferences derived from the environment the individual chooses to surround themselves with (Merleau-Ponty, 2004). Thus, an environmental psychology influences how preferences for certain environments are driven by a multitude of interwoven factors. Jane Jacobs recognised this as early as 1960, emphasising the role streets perform in setting the visual scene of cities. In a critique of modernist planning policy, Jacobs (1961) argued that unifying design elements of urban spaces is short-sighted, as the interplay of their "bits and pieces" are central to supporting the diverse excitement that street scenes offer.

Visual cues are seen as discriminative features that influence perceptions and evaluations of urban spaces, and even when considering socio-cultural biases in aesthetic judgement, have been shown to affect the psychological state of their inhabitants (Quercia et al., 2014). Kelling and Coles (1997)'s *Broken Windows Theory*, for example, suggests cues of environmental disorder in urban appearance such as abandoned cars, litter, and vandalism drive a perceived breakdown of social order which, in turn, induce more severe forms of criminal activity. Beyond disorder places deviate from conventional form by appearing, amongst other things, transgressive, glamorous, or informal (Silver and Clark, 2016). Thus, a suite of evaluative dimensions are considered when characterising the visual attributes of urban spaces, with different environments reflecting different visual representations of tastes and values. Not only this, Massey (1991) argues these particular spaces are not static, but have multiple identities that are forged by ever-changing social interactions occurring between people within them. All together, these considerations highlight the complexities of capturing a signal that reflects the visual qualities of street scenes.

2.2. Traditional approaches for describing retail environments

As aesthetic descriptions of urban environments such as glamorous, lively or conventional are difficult to measure directly, urban scientists typically fall back to constructing indicators of the qualities that describe spaces such as shopping areas (Silver and Clark, 2016). In-person visual audits strive to unpack how the functional, physical and social characteristics of retail environments correlate to affective outcomes such as store patronage and location choice decisions. Survey techniques have an extensive history in urban planning research, and borrow from psychometric measurement models to infer latent traits through an aggregation of single items visible across the audit (Bader et al., 2017). In UK planning discourse, for example, concepts such as vitality and viability have long underlined 'health checks' of town centre areas, reflecting arguments in Jacobs (1961) that thriving places maintain a diverse range of uses, attract significant numbers of people, and sustain a continuing ability to attract investment (Ravenscroft, 2000). Thus, vitality and viability is typically inferred by aggregating multiple items such as pedestrian counts, diversity of amenities, or boarded-up windows that are sampled at points across different retail locations. In the retail literature, several examples aggregate sets of measures to describe visual characteristics of shopping spaces. Bell (1999), for example, shows environmental stimuli such as appealing store colours, attractive shop signs and fashionable product ranges constitute a 'visual amenity' that inspires consumer willingness to patronise a shopping environment. Moreover, El-Adly (2007), finds attractiveness attributes of shopping malls such as luxury, comfort and convenience drive different patronage motives amongst different shopper segments in UAE.

Survey-based approaches are often required to describe the visual properties of urban environments due to the absence of accessible and high coverage quantitative data (Salesses et al., 2013). Traditionally, studies are undertaken by relying on a mix of personal interviews, street-level observations of visual appearances, and annotated video recordings by experts (Quercia et al., 2014). This manual review of material is an arduous task however, and requires considerable collective effort to distinguish amongst the variety of visual cues encoded in the images.

2.3. Deep learning approaches for describing urban environments

To evaluate visual characteristics of particular places, Convolutional Neural Networks (CNNs) that are 'trained' with human-labelled images of street scenes are increasingly used to automate the classification of the scenes presented by built environments. This new body of literature has been punctuated by emerging access to new sources of data that have been released by commercial providers and photo-sharing websites in open formats (Arribas-Bel, 2014). Providers such as Google Street View (GSV) and Flickr have opened up access to street-level imagery for researchers through Application Programming Interfaces (APIs), which have, in turn, been used to construct modern crowdsourcing platforms for collecting millions of user perceptions about particular places. Large quantities of human-labelled, street-level imagery have been used for training computer vision techniques. Zhang et al. (2018), for example, use a deep learning based approach to predict perceptions of neighbourhoods in Bejing, China along six perceptual indicators of safe, lively, boring, wealthy, depressing, and beautiful, before investigating which visual elements correlate to a particular perception. The study used street-level images collated by MIT Media Lab as part of the "Place Pulse" program, which by fall 2018 had collected 1,566,218 pairwise comparisons between 110,988 street-level images from 56 cities worldwide (Dubey et al., 2016). This crowdsourced data was made publicly available by Salesses et al. (2013), who originally used it to understand the effect of the built environment's visual features on perceptions of safety, class and uniqueness in the cities of New York and Boston in the United States, and Linz and Salzburg in Austria. Additional studies that use labelled GSV images include Liu et al., (2016), who detect shifts in city identities and urban form for 26 cities from Europe, Asia, and North America. Seresinhe et al., (2017) trained machine learning models on 217,000 crowdsourced images from the "Scenic-Or-Not" online game that rates outdoor, natural environments on an integer scale (1–10) of its *scenicness*, and explores questions that ask which types of greenspaces are perceived as beautiful.

Unfortunately, a drawback of these supervised methods are the large sample sizes required to train the network which are often unfulfilled in real-life applications. Moreover, these approaches typically utilise a large, non-expert workforce (voting on crowdsourcing platforms) to construct massive volumes of labelled image data. This creates several challenges. Principal amongst these is the balancing between maintaining a swift and economical annotation process while ensuring the collected labels are accurate (Sorokin and Forsyth, 2008). More importantly, the user's interaction with the labelling task may be influenced by socio-economic and demographic factors. As urban experiences are highly socially constructed, different groups might engage with the built environment in different ways, meaning visual characteristics are highly particular to various socio-economic or demographic groups (Quercia et al., 2014). These challenges exist because CNNs are supervised, meaning they require the network to be shown labelled instances of images for learning the nuances between particular predicted outputs. An alternative approach to extracting features from street-level imagery are *Convolutional Autoencoders* (CAEs). CAEs are *unsupervised* approaches meaning they provide a self-organised means for learning the relationships between elements in the data without being shown labelled inputs. CAEs are advantageous because they provide a less data-intensive alternative to CNNs that does not require the user to assemble large quantities of labelled data for training the network.

2.4. Application of computer vision methods to retail environments

While many studies that apply deep learning have focussed on urban environments, to our best knowledge, no application of deep learning to explore visual characteristics of retail environments currently exists in the literature. This is despite the high suitability of computer vision methods for characterising the variance in image attributes between different shopping areas. Consumers with little experience of a store or environment may use perceptual qualifications of image, in addition to prices, as a proxy for the quality of goods and service provision (Bell, 1999). Stimuli that influence consumer perceptions of shopping area image are functional qualities but also the aura of psychological attributes aroused by the environment. Functional characteristics include convenience and accessibility of store or retail area location, parking availability, the range of stores and products offered, and proximity to residential neighbourhoods and workplaces (Baker et al., 1994; Chebat et al., 2010). Psychological characteristics relate to the "visual amenity" experienced by consumers in shopping environments. For example, previous research links store patronage decisions to visual elements such as architecture, shop signage and exterior design (Baker et al., 1994), but also factors such as cleanliness and even colour of store premises (Bellizzi et al., 1983). Thus, quality inference for shopping areas is a function of multiple influences that affect consumer decision-making choices.

Given the wealth of research that has already linked image attributes of shopping areas to consumer patronage, the focus of the present study moves away from an exploration of footfall. Instead, our main research direction focuses on characterising the different *visual* representations of shopping environments by functional attributes that describe the area in which the premise is located. In synthesis of these two attributes, we unpack different representations of the *scene* that particular shopping environments project to passers by. The "*scene*" of an environment reflects both the visual characteristics and configuration of leisure, services, retail and cultural life, and data describing amenities such as leisure and retail premises are windows that allow researchers to unpack these configurations (Silver and Clark, 2016). An understanding of different *scenes* from leisure and retail environments is an important exercise because it unpacks patterns of urban human activity and function. This is useful information for retail planners and urban management schemes because it raises awareness of attributes and image among particular areas. Public or private sector agencies might utilise this to rationalise investment decisions that allocate spend to promotional activities and place marketing campaigns for building the profile of shopping environments (Page and Hardyman, 1996).

The visual design of retail environments are among the tools used to enrich the consumer shopping experience. Visual design of shopping areas has been manipulated previously to evoke desirable responses, such as arousal and pleasure which triggers approach behaviour and supports store positioning (Ballantine et al., 2010; Baker et al., 1994). Yet the visual design of retail environments in the UK is highly particular, and so consideration to its nuances is required for understanding potential implications to our applied methods. One limitation in applying computer vision methods to UK high street environments is a phenomena known as *clone towns* (Ryan-Collins et al., 2010). The idea of 'cloned' streets relates to the loss of identity and local character when chain stores come to homogenise high street environments at the expense of independent stores (Carmona, 2015). The implications for computer vision approaches concern the difficulty in identifying different typologies where no unique characteristics are directly observable from the images when they broadcast no local distinctiveness. Despite the British Retail Consortium (2009) arguing there have been calls for communities to reclaim their local high streets through the encouragement of local spending, it remains that a large number of distinctive facades constructed from local building materials may have been exchanged by identical glass, steel and concrete frontages (Ryan-Collins et al., 2010). This potentially limits the discovery of more interesting, diverse and distinctive types derived from empirical exercises that use street-level imagery from UK high streets. Despite this limitation, for wider study areas than would be permitted by in-person audits, computer vision approaches allow us to unpack how visual features of leisure and retail properties relate to functional characteristics of shopping environments, and consequently, how we can characterise the scenes these places offer.

3. Empirical strategy

Our approach to explore differentiation between visual features of leisure and retail premises is three-staged. Firstly, we extract visual features from images of leisure and retail premises using a computer vision algorithm. Secondly, we partition visual features into a sensible number of clusters using a bottom-up classification strategy. And thirdly, to differentiate between the clusters, we introduce variables that describe characteristics derived from the point of interest around the properties.

3.1. Data

To implement the methodological approach we require two principal sources of data described below. Our first source of data are street-level imagery of 314,542 retail, service and leisure properties across England and Wales. These images display the front exterior of the property that face onto the adjacent street or open space. Exterior images were collected by a large pool of surveying teams equipped with hand-held cameras from the Local Data Company (LDC) in 2015. Sample images are displayed in Fig. 3.1, and are categorised row-wise by several variables introduced in Table 3.1. As a pre-processing step, each JPEG image is resized from $800 \times 400 \times 3$ to a $224 \times 224 \times 3$ pixel

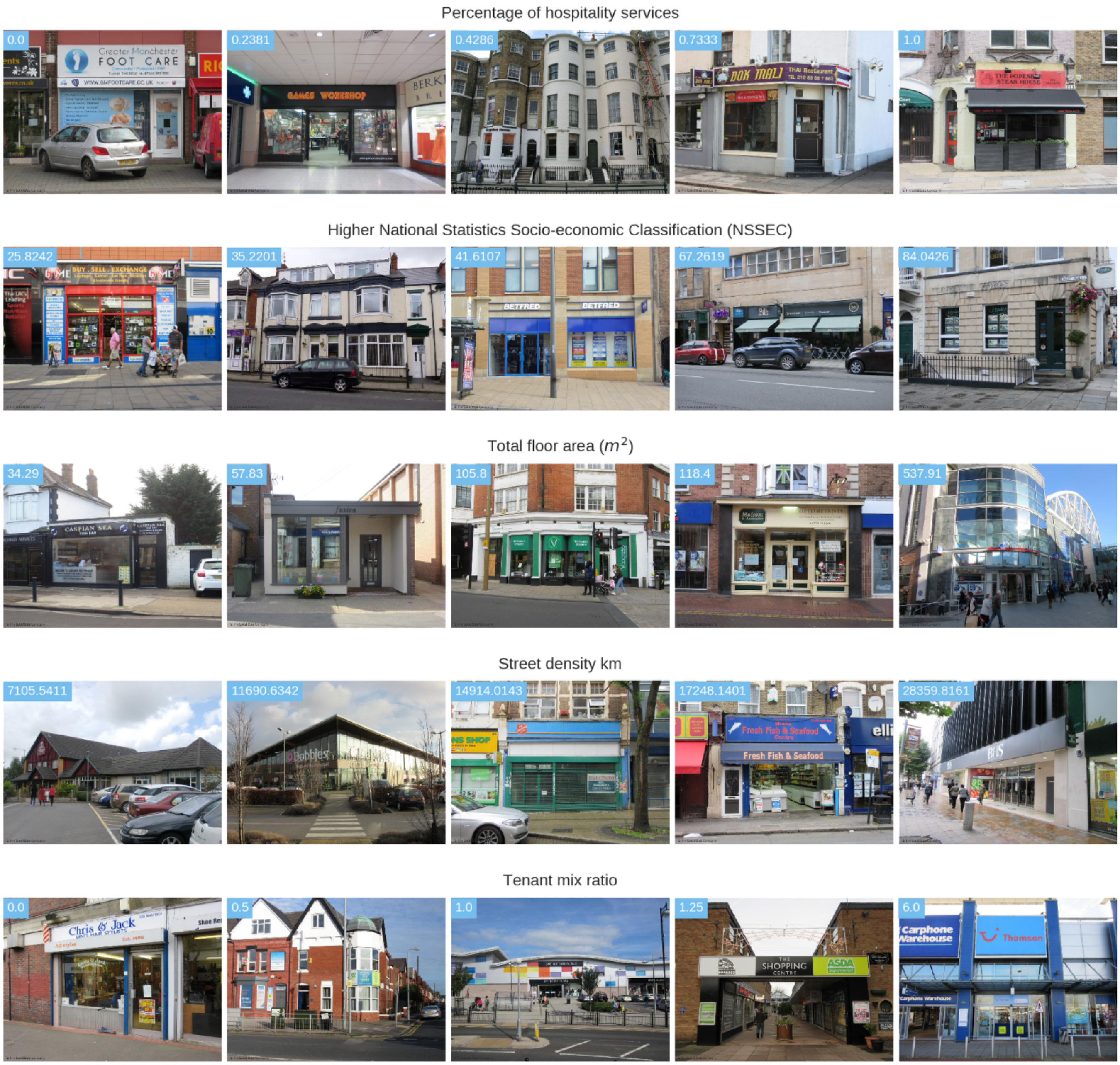


Fig. 3.1. Sample LDC images for several features in Table 3.1. Each image is a random sample from each of five equal interval bins.

image for compatibility with the applied neural network architectures, before normalizing the RGB values (0–255) to a 0–1 range. These resized, normalised digital images are the 3-dimensional inputs (width, height, and colour channel) to the convolutional neural networks we introduce in Section 3.2.

While this data offers new opportunities, there are limitations of using street-level imagery for visual audit purposes. Channels that affect perceptions of built environments such as sound and smell are absent from pictographic representations, and so cannot be directly evaluated from the image (Salesses et al., 2013). Similarly, small items less visible to the human eye that vary over short periods such as litter, drug paraphernalia, broken glass, or cracked sidewalks are difficult to measure given street-level imagery represent a single snapshot in time (Bader et al., 2017). More specifically, given the principle concern for the LDC surveying teams was to photograph facade features of the store premises, measures related to sidewalks such as number of parked cars or shrubbery might be partially occluded in the image, despite contributing to the overall ambiance of the urban area. Despite these limitations, the LDC images remain a valid source of data for our purposes. This is because they simulate a virtual walk down the street that replicates an eye-level experience, and the large number of LDC images provides granular, unprecedented coverage that would be impractical (and cost-intensive) to obtain otherwise.

The second source of data is derived from characteristics that differentiate the particular visual representations of LDC images, and is used in the third stage of our approach. Our variable selection covers measures derived within a 15-minute walk catchment (assuming a walk speed of 4.5 km per hour) around each leisure and retail premise (see Fig. 3.2). These catchments are constructed using OSMnx, which is a Python library for acquiring, analysing and visualising street networks (Boeing, 2017). Within each catchment, we derive measures for a number domains outlined in Dolega et al. (2019) that describe shopping activity such as composition, diversity, size and function, and economic health (see Table 3.1). Aside from LDC and OSMnx data, we derive

Table 3.1
Variable description for the domains of economic health, composition, size and function and socio-economics of leisure and retail premises.

Variable	Description	Source	Mean	Std. Dev	Unit
Economic health					
bus_rate	Rateable value taxed on the business property.	LDC	100,639.9	976,771.5	Pounds
vac_rate	Vacancy rate of Local Authority District the property resides in.	LDC	0.09	0.03	Percent
unemployed	Percent of unemployed people in Output Area.	ONS	5.75	3.64	Percent
e-res_score	E-resilience score of nearest town centre.	CDRC	0.08	0.45	Score
transport	Number of bus or train links within catchment	NaPTAN	61.21	38.19	Count
Composition					
comparison	Proportion of comparison goods stores within catchment (clothing, household goods, etc).	LDC	0.21	0.21	%
hospitality	Proportion of hospitality outlets within catchment (restaurants, bars, etc).	LDC	0.31	0.24	%
convenience	Proportion of food retailers within catchment (grocers, butchers etc).	LDC	0.13	0.18	%
consumer	Proportion of consumer services within catchment (banks, estate agents, etc).	LDC	0.18	0.21	%
tenant_mix	Retail to service ratio of catchment.	LDC	0.88	1.00	Ratio
store_diversity	Diversity of store types within catchment calculated by Shannon entropy.	LDC	1.14	0.57	Bit
Size and function					
floor_area	Total floor area for the property.	LDC	227.61	830.08	m^2
car_parking_spaces	Number of car parking spaces at the property.	LDC	1.28	17.36	Count
roeck_compactness	Compactness of catchment morphology.	OSMnx	0.49	0.13	Ratio
store_diversity	Number of stores within catchment.	LDC	14.24	20.45	Count
eig_centrality	Influence of store location within street network of catchment.	OSMnx	0.02	0.04	Score
street_length_avg	Average length of streets in catchment.	OSMnx	66.43	25.86	Meter
street_density	Total street length within catchment divided by catchment area.	OSMnx	15911.72	7134.42	km^2
Socio-economic					
high_nssec	Percent of people with higher occupational employment in Output Area.	ONS	42.86	17.21	Percent
detached	Percent of housing units classified as detached in Output Area.	ONS	6.09	9.52	Percent
flats	Percent of housing units classified as flats in Output Area.	ONS	35.13	24.94	Percent

variables from several other sources. Census data is provided by the (ONS, 2016), our *e-res_score* variable is from a Consumer Data Research Centre (CDRC) data product and describes the vulnerability of town centres to the impacts of online shopping (estimated by Singleton et al., (2016)), and the *transport* variable is from the database of National Public Transport Access Nodes (NapTAN) (Department for Transport, 2014). In addition, we use a small number of census-based socio-economic characteristics at Output Area (OA) level to describe the area in which the leisure or retail premise resides. OAs are built from postcode units and are the smallest statistical unit for which UK census data is published (ONS, 2019).

3.2. Visual features from CAEs

Given the collection of leisure and retail property images are unlabelled and represented by a large number of raw pixels, a mathematical technique is required to decompose this larger set of correlated variables (or pixels) to a condensed set that captures the most salient characteristics of the image (Efron and Hastie, 2016). To learn this compressed set of variables from the raw pixels we rely on Convolutional Autoencoders (CAEs) (Goodfellow et al., 2016) which are composed of two layers: an encoder layer f_E and a decoder layer f_D. From a non-technical standpoint, the objective of CAEs is to take an input image, I, and reconstruct it as a copy, $\hat{I}$. Internally, CAEs use a hidden layer h that describes a code to reconstruct the image (Goodfellow et al., 2016). This lower dimensional mapping forces the CAE to prioritise aspects of the image that are the most useful for reconstructing a copy from the input image, meaning h learns the most useful properties of the data while discarding redundancies.

CAEs are extensions of autoencoders, which are techniques that essentially reduce the data under consideration to a smaller set of principal values. Practical applications of autoencoders include data compression for saving storage space and transmission times, and also cleaning corrupted data inputs by denoising. Thus, CAEs are autoencoders that introduce convolutional and (de)convolutional layers in the encoder f_E and decoder f_D sections, respectively:

$$f_E = \sigma(I * K + b) = h$$

where σ is a Rectified Linear Unit (Relu) activation function which is a truncation performed individually for every pixel x of the input, $Relu(x_{ij}) = max(0, x_{ij})$, that allows the CAE to learn non-linear patterns in the data, I are $224 \times 224 \times 3$ images where the 3 refers to the red, blue and green (RGB) colour channels, K are 3×3 matrices called convolutional filters, b is the bias unit which is similar to the intercept of a linear function and allows the line of the activation function to shift from the origin, and h is the code that represents the lower dimensional mapping of I. The convolution operator, $I * K$, is described more explicitly for the first layer in Eq. 3.2:

$$(I * K)_{xy} = \sum_{i=1}^{224} \sum_{k=1}^{224} K_{ij} \cdot I_{x+i-1,y+j-1}$$

which overlays each 3×3 filter over every possible pixel of the image, and records the sum of the element-wise product to an intermediate representation known as an activation map. The convolutional operator exploits spatial location in the image, as neighbouring pixels become activated for particular groups of edges that respond to semantically meaningful objects – trees, cars, or people, for example. This means particular filters become activated for specific patterns in the image, and stacking these filters across successive convolutional layers facilitates *parameter sharing*, where hierarchies of filters introduce levels of abstraction to the different kinds of features identified in the image (Goodfellow et al., 2016). As an example, the banks of filters learnt at the first convolutional layer might represent lower-level features such as lines, circles, and curves, while the higher-level convolutional layers will use these to construct whole objects – eye-like shapes or automobile wheels, for example. As the starting values of the K filters are randomly initialized, over the course of training the CAE the network will learn to find the optimal filter values that minimize the reconstruction error between I and $\hat{I}$.

Within each convolutional layer, a final step commonly applied to modify the output from Eq. 3.2 is pooling. We apply the max *pooling* operation which returns the maximum pixel value within a 2×2 filter that steps across non-overlapping pixels of the input. This has the net effect of down-sampling an image by a factor of two, which sequentially reduces the pixel representation of our image from $224 \times 224 \times 3$ to a *latent* representation, h, which has shape $28 \times 28 \times 1$ and reflects the

Fig. 3.2. Example 15-minute walk catchment for a retail store around London Bridge. Note: 30 leisure or retail premises are sampled within the catchment to avoid clutter. Large red star denotes the store for which the catchment was created. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

visual features we use for our clustering exercise (see Section 3.3).

To train the CAE end-to-end, we also require a decoder f_D network that reconstructs the original image $\hat{I}$ from h:

$$f_D = \sigma(h * U + b) = \hat{I}$$

The only difference between f_E and f_D is that convolutional layers in the former are replaced by deconvolutional layers in the latter. This has the net effect of up-sampling the latent representation h ($28 \times 28 \times 1$) back to $224 \times 224 \times 3$, thus completing the reconstruction of the original image I. Once the CAE network has been sufficiently trained, the latent representation h, represented by $28 \times 28 \times 1 = 784$ pixels, becomes the basis of the visual features we use to differentiate between the visual scenes of different leisure and retail premises. To summarise these methodological steps, we visualise the resulting CAE architecture defined by Eq. 3.1 and Eq. 3.2 in Fig. 3.3. In regards to implementation, the CAE model is defined in Keras (Chollet, 2015), with training undertaken on a single Nvidia Quadro M4000 GPU with 8 GB memory.

3.3. Clustering visual features

To derive meaning from the visual features, we require a technique to group our vectors of visual features such that those in the same grouping exhibit similarities. This allows us to unpack similarities between the visual scenes for different retail environments which we can then describe by a number of functional characteristics outlined in Table 3.1. Our approach constructs a *bottom-up* classification where an initial typology with 250 numerous smaller groups are partitioned using k-means. Given the sensitivity of k-means to the initial starting values of the centroids, the algorithm is initialized 1000 times with different centroid seeds, taking the final result as the output that best minimizes the within-cluster sum of squares. Finally, we allow up to 100,000 iterations within a single run to ensure stable convergence of the centroids. After the initial partition, we aggregate the clusters into coarser and larger groupings based Ward's method of hierarchical clustering (Ward, 1963). As Ward's method produces a dendrogram, we use it to slice a horizontal cut along the y-axis to create coarser levels of classification, which groups the 250 centroids of visual features into a smaller number of distinct clusters. This final partition represents the resulting clusters that differentiate the visual characteristics of the LDC images. Thus, we replicate a work flow similar to Spielman and Singleton (2015) and follow simple and widely supported methods to facilitate methodological transparency and reproducibility.

4. Results

In this section, we develop a discussion of our empirical findings based on two validation procedures. First, we undertake a validation exercise on our bottom-up clustering solution to ascertain a desirable

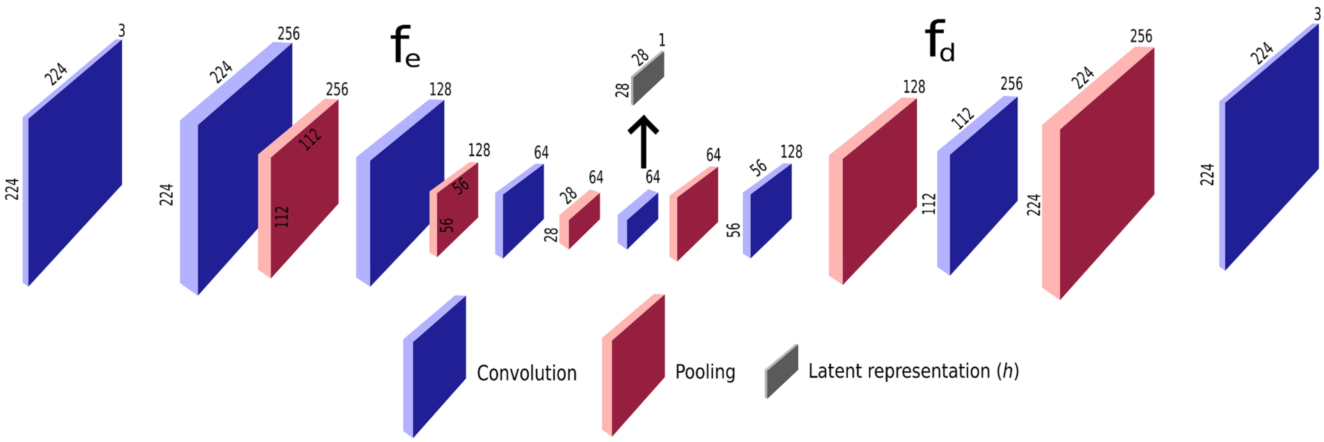


Fig. 3.3. Convolutional Autoencoder (CAE) architecture showing encoder f_e, compressed representation h, decoder f_d and reconstructed LDC image $\hat{I}$. Note: filter numbers are shown horizontally along z-axis of feature maps, while width and height are shown along the x and y, respectively. Illustration was produced on the open-source vector graphics editor Inkscape (Inkscape Project, 2019).

number of clusters; and second, we explore consistency of group membership to particular clusters across sets of visual features generated from the CAE and two pre-trained CNNs. For brevity, the detailed outcome of these exercises are moved to Appendix A and Appendix B. Based on the outcome of these exercises, in the following section we introduce several characteristics to unpack differences between the *five* distinct clusters of images we retrieve from our clustering approach.

4.1. Differentiating visual characteristics

To describe differences between the visual clusters, we aggregate characteristics for the consumer properties from Table 3.1, taking the median value for each variable per cluster. Prior to the aggregation, we transform each variable to z-scores by standardization, $z = \frac{x-\mu}{\sigma}$, meaning each characteristic is rescaled by the fractional number of standard deviations from the mean value. To begin, we introduce radar plots in Fig. 4.1 where each plot reflects a different visual cluster that shares similar psychological attributes reflected by common visual elements such as similar exterior design, signage, architecture, or colour. Along the axis of each plot aggregated variables that describe functional characteristics of these clusters are displayed. Thus, in synthesis of visual (psychological) attributes revealed by the cluster groupings and functional characteristics by the variables, we describe the *scene* projected by the clusters.

Turning to the group sizes, we note the numbers of leisure and retail premises within the visual clusters vary substantially. Our largest cluster, Group A, contains 159,251 leisure and retail properties whose built environment is distinguished by high density street networks and large proportions of comparison retail outlets who sell merchandise that consumers purchase relatively infrequently and so evaluate prices, features and quality between stores before making a purchase. This includes outlets such as DIY & household goods, electrical, and clothing and footwear stores. Group A also contains a considerable proportion of hospitality outlets such as restaurants, bars and pubs, and entertainment venues. The Roeck compactness value measures irregularity in the shape of the retail area's boundary, with higher values indicating a highly compact retail area and lower values reflecting dispersion. The Roeck value for Group A, alongside its high street density, implies the urban morphology of the built environment around these stores is highly dense and not dispersed. All together, this suggests the scene characteristics of Group A reflects a bustling shopping area with relatively affluent residents who live in the immediate area (as shown by the high percentage of residents in higher occupational roles).

Group B contains 24,567 leisure and retail premises and is highly differentiated amongst its characteristics when compared to the other clusters. The functional attributes shared by leisure and retail premises inside this visual grouping reflect areas that have a low diversity of premise types, with the majority of outlets represented by comparison retail or consumer services such as car showrooms and house & home stores. Premises in this cluster are located in areas with high vacancy rates, meaning there are higher percentages of vacant or unoccupied store units relative to the other groupings. Moreover, outlets in this cluster appear to have high total floor areas and are serviced by fewer transport options, which conjures images of peri-urban spaces consisting of large retail units and warehouse spaces located on the fringes of dense urban areas and so are less beaming with consumer activity. Overall, the visual and functional characteristics of Group B portray a scene of sparse and less desirable retail and leisure land use when compared with the other clusters. This is reinforced by socio-economic characteristics which reveal that individuals who live in the area, and might patron the shopping environment as consumers, typically occupy low percentages of high paid employment.

The next grouping that shares visual similarity is Group C, which contains 81,310 leisure and retail premises and is ascribed the label of 'Upmarket Hospitality'. The shopping environment of premises in this cluster are reflected by a large proportion of diverse hospitality outlets and leisure venues. This includes services ranging from restaurants and bars to theatres and galleries. A second defining characteristic of Group C is the extremely low vacancy rate when compared with the other clusters. This shows store units around the built environment for this grouping are typically occupied, which implies units in this cluster are in higher demand and so possibly elicit increased rates of rent. Similar to Group A, catchments around premises in this cluster are well served by transport links and possess highly similar urban morphology and socio-economic characteristics. In synthesis of visual similarities for leisure and retail premises within the cluster and functional characteristics of the urban landscape around these premises, Group C projects the scene of a thriving and upmarket shopping environment that is highly accessible and amenable to consumption activity.

Our smallest grouping, Group D, contains 6,962 leisure and retail units and is highly similar to Group C, although there are a few variables that differentiate the two clusters. Like Group C, Group D is characterised by a diverse range of hospitality outlets and stores that provide comparison goods such as electrical appliances and clothing. Compared to the dense street network of Group C, the urban morphology of Group D appears to reflect longer average street lengths that

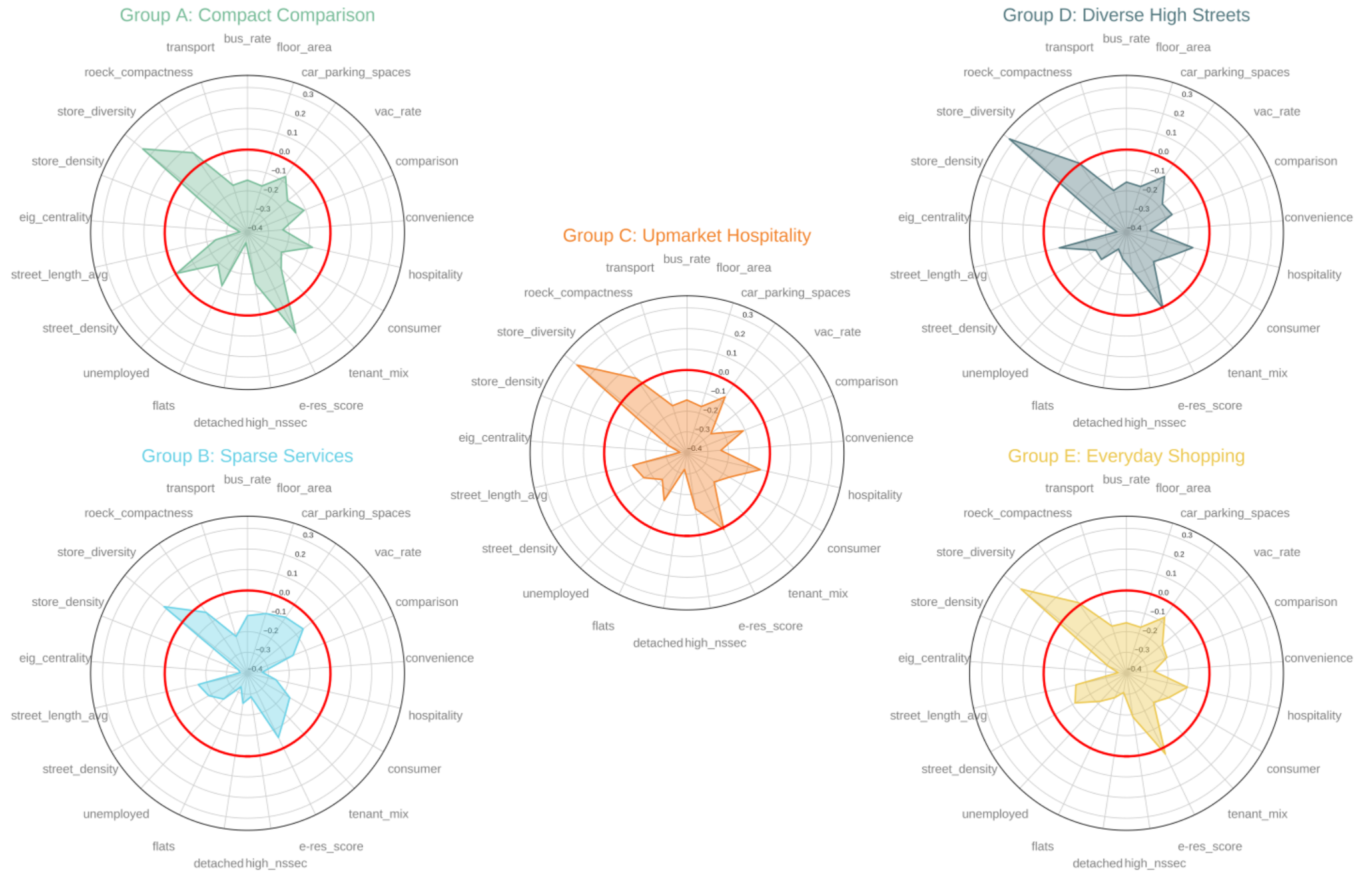


Fig. 4.1. Median economic health, composition, size and function, and socio-economic characteristics in standardized units. Circular red line identifies zero, which shows standard deviations from the mean value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are fairly dispersed as shown by the low street density. Consistent with conventional wisdom, these two observations imply the built environment surrounding leisure and retail premises of Group D reflects high street shopping areas. Residents who occupy residential housing near stores in Group D typically occupy lower proportions of higher managerial roles. This suggests consumers, and by extension local consumption opportunities, are represented by less upmarket leisure and retail outlets given local patrons are typically less affluent than in Group C. Nine example images comparing low to high average street length for Group A and D, respectively, are shown by Fig. 4.2. The presence of automobiles in images sampled from Group D suggest the built environment here is more amenable to vehicle use, with streets around leisure and retail premises in this cluster typically longer and less dense. All together, the composite visual and functional characteristics of Group D project a scene of long high streets that serve a diverse range of consumption purposes to local consumers.

The last cluster, Group E, contains 81,310 leisure and retail premises and represents a middle ground between Group C and Group E. While units providing hospitality represent the highest proportion of services in this cluster, no particular mode of retail or leisure dominates unlike the other groupings. In fact, premises in Group E have the lowest proportion of comparison retailers in the surrounding urban environment. The urban morphology of Group E is fairly dense and compact, as evidenced by a relatively high street network density and Roeck compactness value. In synthesis, the shared functional attributes of premises in Group E suggest this grouping reflects a leisure, services and shopping environment that is accessed by consumers for everyday consumption as opposed to being accessed for a particular mode of retail or leisure service.

5. Discussion and conclusions

Visual characteristics of shopping environments are a significant determinant of area consideration and choice (Bell, 1999). Traditionally, visual representations of retail areas are retrieved using teams of human surveyors, who are cost-intensive to train and limited in the throughput necessary to construct the visual form of built environments. Consequently, in this paper, we use vast quantities of street-level imagery to explore whether visual features of leisure and retail environments correlate to measurable characteristics of built environments. This was achieved using a deep learning model known as Convolutional Autoencoders (CAEs) which learnt a compressed representation that captured the most salient characteristics required to reconstruct the image from a lower dimensional representation. Once these visual features were partitioned into a sensible number of clusters, functional characteristics that describe a 15-minute walk catchment from each premise were introduced to differentiate between the cluster partitions. By clustering the compressed representation, we were able to identify five partitions from the data that reflected different categorisations of the *scene* that particular shopping environments project to consumers across a national extent. This is important because information describing retail area image has historically been desired by retail planners for rationalising investment decisions in place marketing campaigns (Page and Hardyman, 1996), but is seldom available at wide geographical scales.

Furthermore, our findings unpacked patterns of retail activity and function, which demonstrated that certain visual features were distinctive for particular built environments. From an urban planning perspective, the main implications of our study demonstrated that

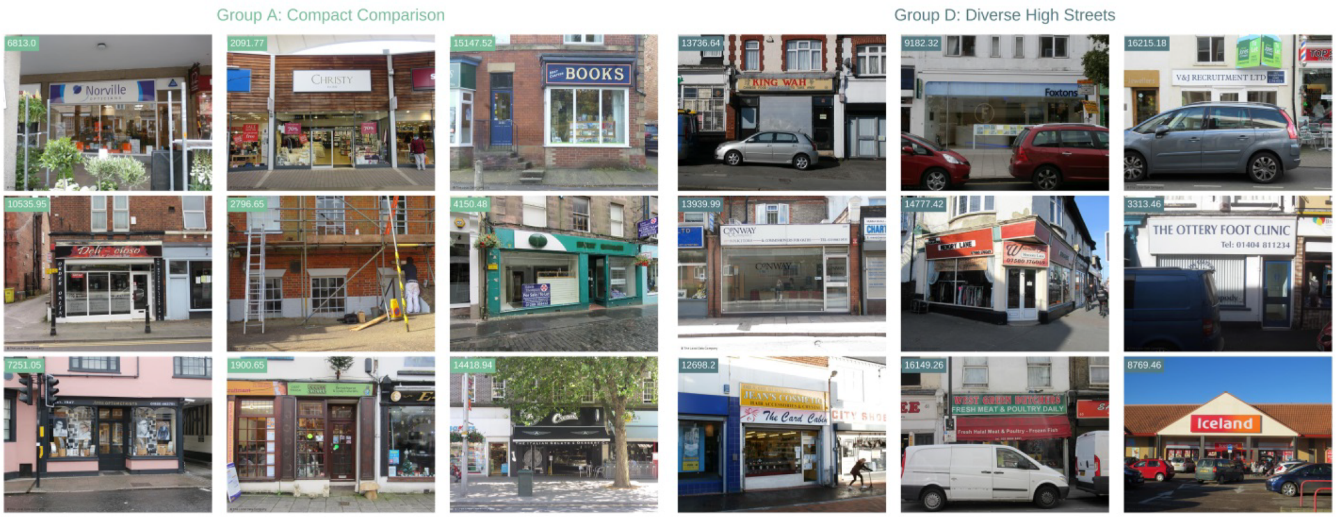


Fig. 4.2. Leisure and retail storefront images and average street length values in metres sampled from Group A and Group D.

aspects of what humans *see* were related to particular functional characteristics of retail environments. This was a pertinent question for retail practitioners to ask, as while previous studies have shown that *proximity* to (and *attractiveness* of) amenities such as leisure plazas, galleries and shops enter into consumer patronage decisions (Glaeser and Gottlieb, 2009), the defining visual characteristics of these environments are typically ignored. This is despite visual amenity being an important influence on patronage behaviour and the *scene* that shopping environments project to consumers (Silver and Clark, 2016).

In more practical terms, our approach could be mobilised within retail planning by adding a visual dimension to retail site optimization tools, and be used to optimally locate stores in locations suitable to particular consumer space uses. More precisely, retail managers could take photographs of prospective site locations, and classify each one according to the several clusters we identify. This would require passing the photographs through the CAE, and using the clustering outcomes fitted on the LDC images to predict cluster membership of these new, unseen photographs. Our approach, therefore, could be used to contextualise the visual qualities of potential store locations among storefronts that look visually similar through observing which particular environmental variables are atypical of the visual cluster this new image belongs to. A retail manger interested in siting a restaurant, for example, could photograph several prospective locations up for sale, and use our approach to retrieve a classification for each. The resulting classification would provide information describing whether the visual qualities of each location reflect typical uses of these spaces that are suited to their business. Following our restaurant example, a photograph classified as sharing visual commonalities to our Upmarket Hospitality cluster would likely present the most idealised location, by highlighting this photograph shares visual similarity to locations that appear to attract high volumes of hospitality services. In using this approach to complement existing tools, we argue taking into account the visual amenity of potential locations could help retail planners to arrive at *smarter* site location decisions, which carries wider implications for the vitality of town centres when amenities within these consumption spaces are optimally situated.

More generally, replication of our approach on a similar corpus of images (Google Street View, for example) could be used by planners to find whether different visual environments reflect particular patterns of built environment use, crime or socio-economic conditions of an area. Across particular urban centres, for example, planners might collect similar image-based data and apply our methods to identify visual commonality between different locations. Then, by collecting a set of variables of interest describing each location, planners might identify similarity or dissimilarity across different variables between the visual clusters. By example, if planners find a particular cluster suffers disproportionately high crime rates, they could sample a number of images from this cluster and undertake post-hoc analysis on possible visual cues embedded in images of these locations. In doing so, our approach provides means for planners to evaluate visual elements that potentially drive the incidence of conditions like crime, which might be identified from locations with high enclosure or no street lighting, for example.

A further contribution of the present study relates to several methodological innovations we introduce in the analysis. As our CAE model is unsupervised, it does not require large numbers of labelled images for training the model to produce visual features for each image. While the existing focus of the literature uses pre-trained or fine-tuned Convolutional Neural Networks (CNNs) for computer vision tasks in urban planning (Dubey et al., 2016; Seresinhe et al., 2017; Zhang et al., 2018), in the present paper we show that unsupervised techniques such as CAEs can also extract visual information from street-level imagery. This is advantageous for two reasons. Firstly, it does not require the user to assemble a large number of labelled images for training the CNN, which might possibly be derived from a non-expect workforce on a crowd-sourcing platform such as Amazon Mechanical Turk. And secondly, because pre-trained networks are often designed for a different purpose than that intended by the user, transfer learning approaches may provide sub-optimal performance if the images used are too heavily skewed compared to the data used to train the original network. Thus, while CNNs can be fine-tuned to the user's image data, a secondary contribution of this paper highlights the utility of CAEs for urban scientific tasks seeking to extract visual information from street-level imagery.

Despite these advantages, there exists conceptual and methodological limitations that frame the conditions for which the study should be interpreted. From a conceptual standpoint, it is reasonable to suggest the 15-minute walk catchment used to derive measures that describe the functional characteristics of the environment around each premise might not be reflective of reality on the ground. A 15-minute walk in a dense urban environment like London is likely to intersect a variety of scenes that possess polarised socio-economic and functional characteristics – for example, the short distance between the affluent and poorer areas of Clapham and Brixton, respectively. This means measures describing the built environment within each catchment might be inaccurate due boundary effects that influence area consideration and create barriers beyond which consumers do not patronize. From a methodological perspective, a further limitation is that repeatability of the empirical approach is conditional on the availability of suitable GPU hardware for training the CAE model end-to-end. Unfortunately, deep learning models require appropriate hardware to train, and this presents a financial barrier of access to researchers interested in replicating (or extending) the empirical strategy to their own datasets. Despite these concerns, the main contribution of this article presents directions for future researchers to employ the deep learning methods adopted by the paper. As CAE networks are unsupervised, they offer flexibility to researchers seeking to extract visual features from image data without using pre-trained networks. This is a pertinent point to consider because the target domains of pre-trained networks are often purposed to answer a different research question than that asked by the user.

Credit authorship contribution statement

Sam Comber: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Daniel Arribas-Bel:** Supervision, Writing - review & editing, Funding acquisition. **Alex Singleton:** Supervision, Writing - review & editing, Funding acquisition. **Les Dolega:** Supervision, Writing - review & editing, Funding acquisition.

Appendix

A. Cluster validation

The lack of a single global optimization procedure is an inherent limitation of clustering exercises, meaning the plausibility and usefulness of the classification are typically split between the purpose it serves but also a validation of its system-wide accuracy. With this in mind, we pair human intuition for ascertaining a sensible number of clusters alongside a metric used for measuring cluster compactness known as average silhouette width. To determine the quality of possible cuts to the dendrogram and, therefore, resulting number of final clusters, we calculate the average silhouette width for several partitions of the 250-class k-means solution. Silhouette width ranges from $-1 \leq s_i \leq 1$, with higher values being desirable as they imply low within-cluster dissimilarity; it is calculated as $s_i = \frac{b_i - a_i}{max(a_i, b_i)}$, where a_i is the average Euclidean distance of i to all other data points in the same cluster, and b_i is the Euclidean distance of i to the cluster nearest to the one i is assigned to.

In practice, we average s_i for all observations for each cut from 2 to 249 of the dendrogram in Fig. A.1, taking the final cut as one that yields a high average silhouette and sensible number of clusters. By scanning the figure we are able to discern a sensible number of five clusters which is ideal because five is both manageable to describe and large enough unpack interesting between-cluster variation. To accompany this, we provide the resulting dendrogram for the five clusters in Fig. A.2, which visualises the agglomerative steps used to aggregate the 250-class k-means solution into five coarser groupings. This is important because hierarchical clustering techniques do not provide cluster partitions automatically, and so *tree-cutting* procedures are required to return partitions that reflect similarities amongst observations in the agglomerative procedure. In our case, while other cuts to the dendrogram offered reasonable performance, we take the decision to cut the dendrogram horizontally at this particular position (of the y-axis in Fig. A.2) because the five cluster solution has a high average silhouette width and sensible number of clusters.

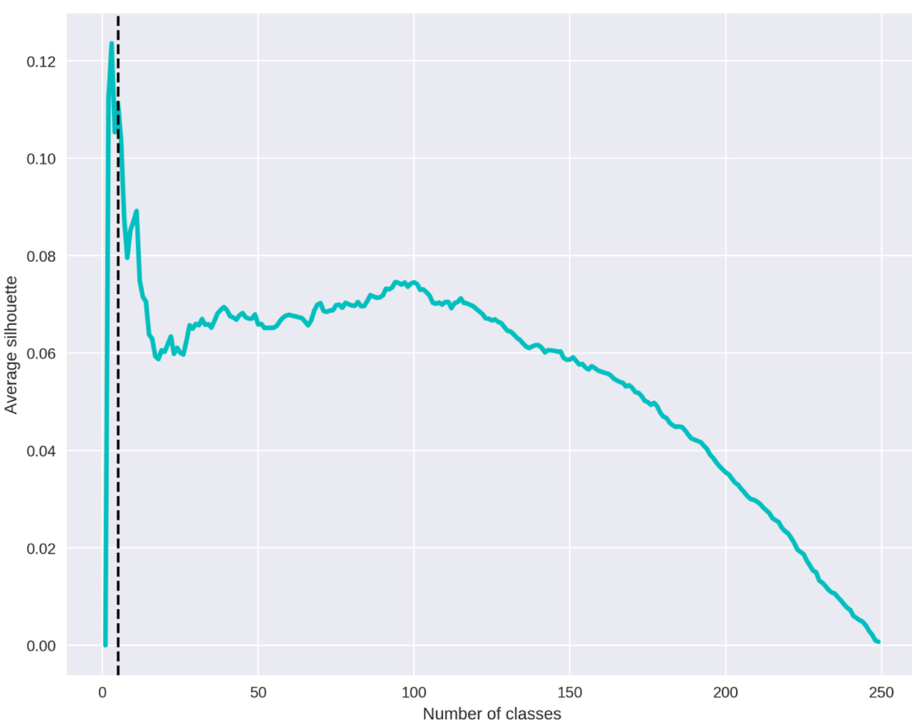


Fig. A1. Average silhouette for different aggregations of the 250-class k-means solution. Vertical dashed line indicates the desired five-class solution.

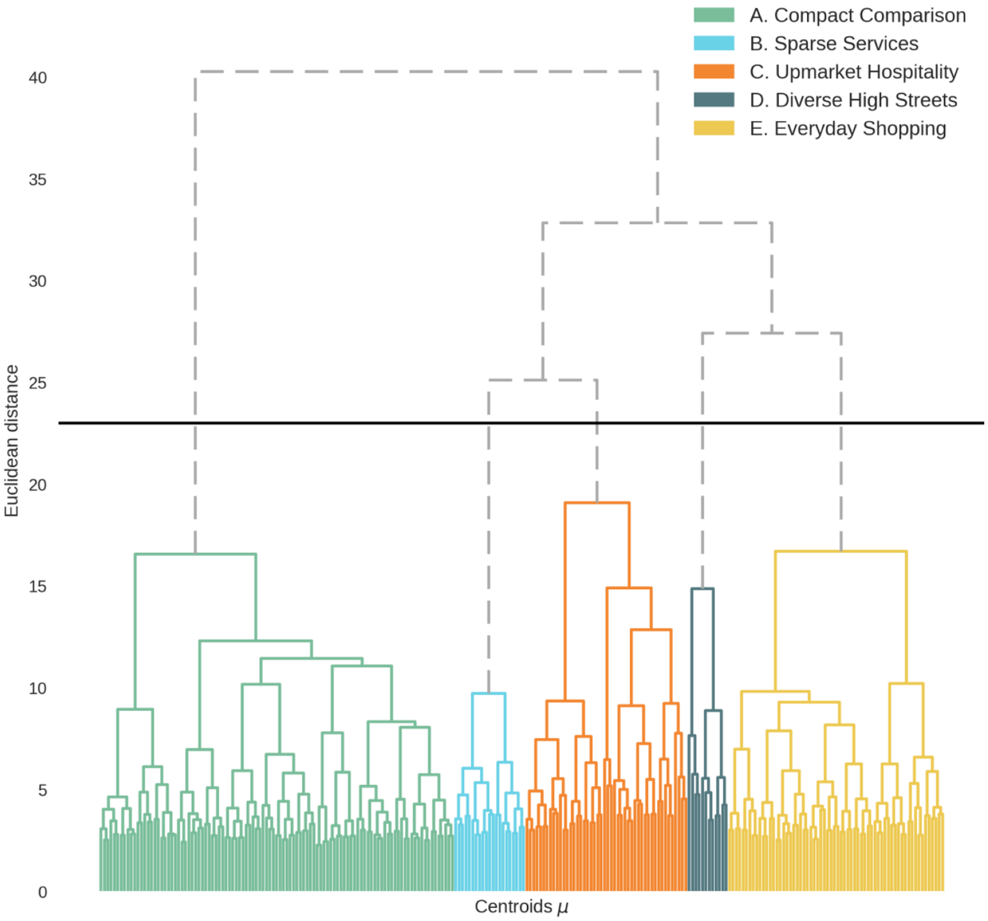


Fig. A2. Dendrogram displaying the agglomerative merge of the 250-class k-means solution.

B. Consistency with pre-trained visual features

To benchmark the visual features, h, retrieved from the latent representation encoded by the CAE we extract a similar set of visual features from two pre-trained Convolutional Neural Networks (CNNs): VGG16-Places365 (Kalliatakis, 2017) and ResNet50 (He et al., 2015). While pre-trained CNNs are trained using large volumes of labelled data for predicting a pre-defined set of categories, CAEs learn visual information that is optimised to the dataset supplied by the researcher. Between these approaches reflects a trade-off between the generalisability of CNNs to extract features learnt from a larger pool of images and more focused visual information extracted from the CAE trained on the researcher's data. Irrespective of this, both serve as points of comparison to assess the consistency of group memberships to particular clusters across different sets of visual features. Given these networks are pre-trained, they are not required to be trained from scratch, and so are initialized with existing weights. For VGG16-Places365, the network weights are initialized to those trained on the Places365 database consisting of 365 different environment categories – highways, vineyards, or libraries, for example – and are tuned for scene recognition tasks. ResNet50, on the other hand, is initialized with weights trained on the ImageNet database, which is a large visual dataset consisting of hand-annotated images that represent a wider range of 20,000 categories. For these pre-trained networks, we remove the fully-connected layer at the top of the network, meaning instead of returning probabilities for categories, we extract the visual features that are discriminative towards particular categories instead. In all, three sets of visual features are introduced to the clustering exercise introduced below. This includes visual features from the CAE represented by 784 pixels, VGG-Places365 features by 512 pixels, and ResNet50 features by 2048 pixels.

To externally validate our empirical approach we monitor changes in group membership and cluster sizes between visual features extracted from our CAE and the two pre-trained convolutional neural networks (CNNs), VGG16-Places365 and ResNet50. Thus, after clustering each set of visual features from the three models, we explore *agreeability* of cluster membership for a five cluster solution in Fig. B.1. The cluster sizes are represented by the vertical white rectangles for the CAE, VGG16-Places365, and ResNet50 models (left to right), with the frequency of leisure and retail amenities changing between groupings shown by the stream fields, and so represent changes in the composition of clusters between the three models. From an initial reading of the figure a mixed picture emerges. While the group sizes are moderately consistent between the CAE and VGG16-Places365, the clusters formed from the visual features of ResNet50 are far more balanced, with leisure and retail amenities spread more equally amongst the partitions. In regards to group membership, the highest agreeability is observable between the largest clusters partitioned using visual features of the CAE and VGG16-Places365 models. Similarly, the clusters identified by '0' in both models seem to share moderate agreeability, with there also being minor agreeability between '2' and '4' of the CAE and VGG16-Places365 models, respectively; the frequency flows of the remaining clusters are far more dispersed between different clustering solutions. Agreeability with ResNet50 visual features, on the other hand, is observably low, with there being no discernible patterns and consistencies between the clustering solutions. This is unsurprising given the target domain of both pre-trained networks is highly dissimilar, a phenomena known as data bias (Chen et al., 2017). While VGG16-Places365 is optimized for scene recognition tasks, ResNet50 is trained to predict over 20,000 object categories from the ImageNet database, with classes ranging from particular types of plants to bedroom items. The weights of the ResNet50 network are tuned to generate visual features that are discriminative for a wider range

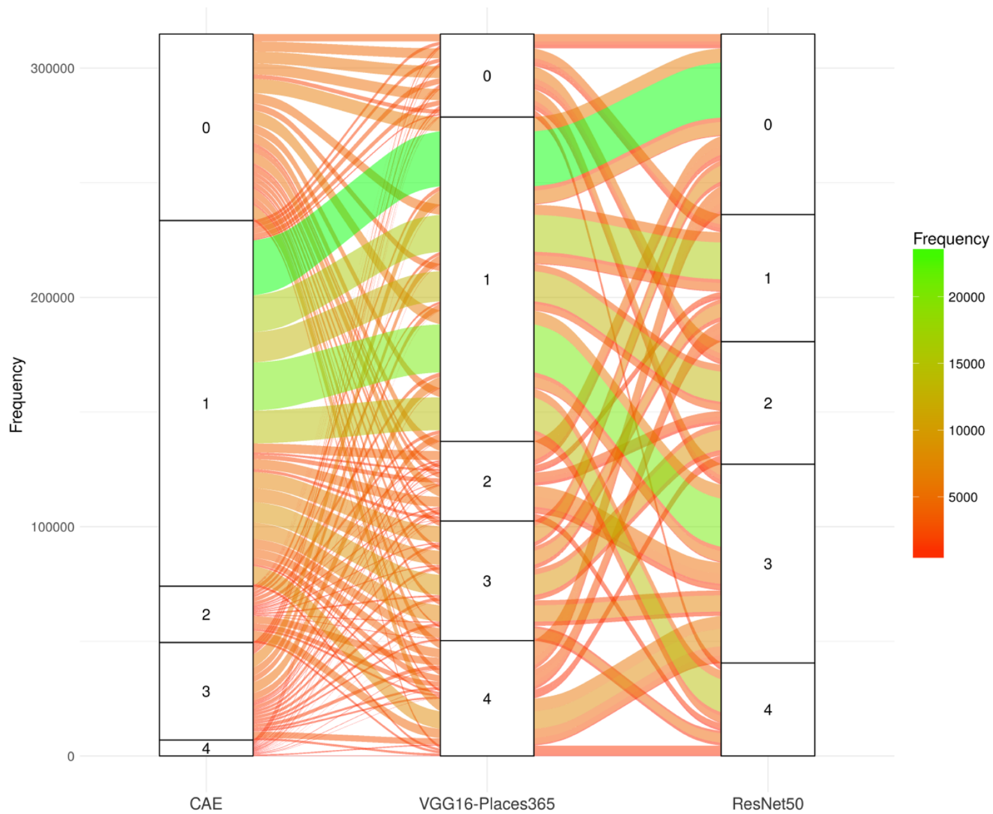


Fig. B1. Agreeability of the fiver cluster solution for visual features from Convolutional Autoencoder (CAE), VGG16-Places365, and ResNet50.

of object classes, meaning when we recover a representation for each leisure or retail amenity image, the kinds of features activated are more generalised than those from VGG16-Places365. This is due to the narrow focus for the range of categories that VGG16-Places365 has been trained to identify (with an emphasis on scene recognition tasks), meaning the visual features are more likely to be similar to those derived from the CAE model. Therefore, as the LDC images describe scenes observable from street-level, there is likely higher agreeability between the CAE and VGG16-Places365 models in terms of group membership and cluster sizes, which is reflected in the figure. All together, these observations confirm the visual features we extract using the CAE model are representing salient properties of the image, which motivates our descriptions for the characteristics of particular visual clusters in our empirical findings section.

References

Arribas-Bel, D. (2014). Accidental, open and everywhere: emerging data sources for the understanding of cities. *Applied Geography, 49*, 45–53.

Bader, M., Mooney, S., Bennett, B., & Rundle, A. (2017). The promise, practicalities, and perils of virtually auditing neighborhoods using google street view. *The Annals of the American Academy of Political and Social Science, 669*(1), 18–40.

Baker, J., Grewal, D., & Parasuraman, A. (1994). The influence of store environment on quality inferences and store image. *J Acad Mark Sci, 22*(September), 328–339.

Ballantine, P., Jack, R., & Parsons, A. (2010). Atmospheric cues and their effect on the hedonic retail experience. *International Journal of Retail & Distribution Management, 38*(June), 641–653. https://doi.org/10.1108/09590551011057453.

Bell, S. (1999). Image and consumer attraction to intraurban retail areas: An environmental psychology approach. *Journal of Retailing and Consumer Services, 6*(2), 67–78.

Bellizzi, J., Crowley, A., & Hasty, R. (1983). The effects of color in store design. *Journal of Retailing, 59*, 21–45.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems, 65*, 126–139.

British Retail Consortium. (2009). "21st Century High Streets: A New Vision for Our Town Centres." http://www.boots-uk.com/media/3866/21st_century_high_streets_2012-1-1.pdf [Accessed 08/11/2019].

Carmona, M. (2015). London's local high streets: The problems, potential and complexities of mixed street corridors. *Progress in Planning, 100*, 1–84.

Chebat, J., Sirgy, J., & Grzeskowiak, S. (2010). How can shopping mall management best capture mall image? *Journal of Business Research, 63*(7), 735–740.

Chen, Y., Chen, W., Chen, Y., Tsai, B., Wang, Y., & Sun, M. (2017). No more discrimination: Cross city adaptation of road scene segmenters. *CoRR, abs/1704.08509*.

Chollet, François. (2015). "Keras." https://keras.io. [Accessed 08/11/2019].

Dennis, C., Marsland, D., & Cockett, T. (2002). Central place practice: Shopping centre attractiveness measures, hinterland boundaries and the Uk retail hierarchy. *Journal of Retailing and Consumer Services, 9*(4), 185–199.

Department for Transport. (2014). "National Public Transport Access Nodes (Naptan)." https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan.

Dolega, L., Reynolds, J., Singleton, A., & Pavlis, M. (2019). Beyond retail: New ways of classifying Uk shopping and consumption spaces. *Environment and Planning B: Urban Analytics and City Science*.

Dubey, A., Naik, N., Parikh, D., Raskar, R., and Hidalgo, C. (2016). "Deep Learning the City : Quantifying Urban Perception at A Global Scale." CoRR abs/1608.01769. http://arxiv.org/abs/1608.01769.

Efron, B. and Hastie, T. (2016). Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. 1st ed. New York, NY, USA: Cambridge University Press.

El-Adly, M. (2007). Shopping malls attractiveness: A segmentation approach. *International Journal of Retail and Distribution Management, 35*(11), 936–950.

Glaeser, E., & Gottlieb, J. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature, 47*(4), 983–1028. https://doi.org/10.1257/jel.47.4.983.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA, USA: MIT Press.

Hauser, J., & Koppelman, F. (1979). Alternative perceptual mapping techniques: relative accuracy and usefulness. *Journal of Marketing Research, 16*(4), 495–506.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Deep Residual Learning for Image Recognition." arXiv Preprint arXiv:1512.03385.

Inkscape Project. (2019). "Inkscape." https://inkscape.org. [Accessed 08/11/2019].

Jacobs, J. (1961). *The death and life of great american cities.* New York, NY, USA: Random house.

Jung, H., Lee, S., Kim, H., & Lee, J. (2017). Does improving the physical street environment create satisfactory and active streets? Evidence from Seoul's Design Street Project. *Transportation Research Part D: Transport and Environment, 50*, 269–279.

Kalliatakis, G. (2017). "Keras-Vgg16-Places365." https://github.com/GKalliatakis/Keras-

VGG16-places365; GitHub. [Accessed 08/11/2019].
Kant, I. (1790). Critique of Judgment. New York City, NY, USA: Barnes & Noble.
Kelling, G., & Coles, C. (1997). *Fixing broken windows: Restoring order and reducing crime in our communities.* A Touchstone Book: Free Press.
Liu, L., Zhou, B., Zhao, J., & Ryan, B. (2016). C-Image: City cognitive mapping through geo-tagged photos. *GeoJournal, 81*(6), 817–861.
Lynch, K. (1960). *The image of the city.* Cambridge, MA, USA: MIT Press.
Massey, D. (1991). A global sense of place. *Marxism Today, 38*, 24–29.
Merleau-Ponty, M. (2004). *The world of perception.* Abingdon, Oxford, UK: Routledge.
ONS. (2016) "National Records of Scotland; Northern Ireland Statistics and Research Agency (2016): 2011 Census Aggregate Data. UK Data Service (Edition: June 2016)." Office for National Statistics. https://census.ukdataservice.ac.uk/use-data/citing-data.aspx. [Accessed 08/11/2019].
-. (2019). "Introduction to Output Areas - the Building Block of Census Geography." 2019. https://www.ons.gov.uk/census/2001censusandearlier/dataandproducts/outputgeography/outputareas. [Accessed 08/11/2019].
Page, S., & Hardyman, R. (1996). Place marketing and town centre management: A new tool for urban revitalization. *Cities, 13*(3), 153–164.
Quercia, D., O'Hare, N., & Cramer, H. (2014). *Aesthetic capital: What makes london look beautiful, quiet, and happy?* (pp. 945–955). CSCW '14. New York, NY, USA: ACM. https://doi.org/10.1145/2531602.2531613.
Ravenscroft, N. (2000). The vitality and viability of town centres. *Urban Studies, 37*(13), 2533–2549.
Ryan-Collins, J., Cox, J., Potts, R., and Squires, P. (2010). "Re-Imagining the High Street - Escape from Clone Town Britain." New Economics Foundation. https://b.3cdn.net/nefoundation/1da089b4b1e66ba2b3_v8m6b0c0w.pdf. [Accessed 08/11/2019].
Salesses, K., Schechtner, K. and Hidalgo, C. (2013). "The Collaborative Image of the City: Mapping the Inequality of Urban Perception." PLOS ONE 8 (7). Public Library of Science:1–12.
Seresinhe, C., Preis, T., Moat, H. (2017). "Using Deep Learning to Quantify the Beauty of Outdoor Places." Royal Society Open Science 4 (7). The Royal Society. https://doi.org/10.1098/rsos.170170.
Silver, D., & Clark, T. (2016). *Scenescapes: How qualities of place shape social life.* Chicago USA: University of Chicago Press.
Singleton, A., Dolega, L., Riddlesden, D., & Longley, P. (2016). Measuring the spatial vulnerability of retail centres to online consumption through a framework of E-resilience. *Geoforum, 69*, 5–18.
Sorokin, A, Forsyth, D. (2008). "Utility Data Annotation with Amazon Mechanical Turk." In 2008 Ieee Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Cvpr Workshops. https://doi.org/10.1109/CVPRW.2008.4562953.
Spielman, S., & Singleton, A. (2015). Studying neighborhoods using uncertain data from the american community survey: A contextual approach. *Annals of the Association of American Geographers, 105*(5), 1003–1025.
Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845.
Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning, 180*, 148–160.